AD-A251 430

**SOLID STATE
ELECTRONICS
LABORATORY**

JSEP ANNUAL REPORT

DTIC
ELECTE
S JUN 8 1992
C D

11 January, 1991 through 10 January, 1992

James S. Harris, Jr.
JSEP Principal Investigator
and Program Director

(415) 723-9775

92-14593

92 6 03 006

STANFORD ELECTRONICS LABORATORIES • DEPARTMENT OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY • STANFORD, CA 94305-4055

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
Unclassified	

2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Stanford University		U. S. Army Research Office

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
Solid State Electronics Laboratory McCullough 226 Stanford, CA 94305	P. O. Box 12211 Research Triangle Park, NC 27709-2211

8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
U. S. Army Research Office		

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
P. O. Box 12211 Research Triangle Park, NC 27709-2211				

11. TITLE (Include Security Classification)

JSEP Annual Report   January 11, 1991 - January 10, 1992

12. PERSONAL AUTHOR(S)
J. S. Harris, Principal Investigator and Program Director

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 1/11/91 TO 1/10/92	5/92	84

16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This is the annual report of the research conducted at the Stanford Electronics Laboratories under the sponsorship of the Joint Services Electronics Program from May 1, 1987 through April 30, 1988. This report summarizes the area of research, identifies the most significant results and lists the dissertations and publications sponsored by the contract DAAG29-85-K-0048.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
J. S. Harris	415-723-9775	

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted. All other editions are obsolete.

JSEP ANNUAL REPORT

Period of 11 January 1991 - 10 January 1992

Department of Electrical Engineering
Stanford University
Stanford, CA 94305

Joint Services Electronics Program
(U.S. Army, U.S. Navy and U.S. Air Force)
Contract DAAL03-91-C-0010

James S. Harris, Jr.
Principal Investigator
and
Program Director

Monitored by U.S. Army Research Office

Abstract

This is the annual report of the research conducted at the Stanford Electronics Laboratories under the sponsorship of the Joint Services Electronics Program from 11 January 1991 through 10 January 1992. This report summarizes the area of research, identifies the most significant results and lists the dissertations and publications sponsored by the contract DAAL03-91-C-0010.

Key Words and Phrases: None

TABLE OF CONTENTS

JSEP ANNUAL REPORT

January 11, 1989 - January 10, 1991

Introduction

The JSEP contract supports a program of unclassified basic research in electronics conducted by faculty members of the Electrical Engineering Department of Stanford University as a component of the research program of the Stanford Electronics Laboratories. The Stanford Electronics Lab JSEP Director and Principal Investigator is Professor James Harris. He is responsible for the selection of the best individual proposals, coordination between Stanford and the JSEP TCC and coordination between the selected areas of the JSEP Program. In planning the JSEP Program at SEL, a general objective is to develop new projects with 3-4 years of JSEP sponsorship, leading to a transition to more conventional DoD or other agency program funding. Since this type of funding often requires 12-18 months in the proposal, evaluation and budgeting stages, the flexibility in JSEP funding allows us to seize new opportunities and initiate programs which might otherwise be delayed for a significant period. An outstanding example of this initiation funding was the Molecular Beam Epitaxial growth of high temperature superconductors, which was supported by JSEP for two years while the DoD high T_c programs were being organized and funded. This project is now supported by DARPA in a large collaborative program with Varian Associates. Highlights of four projects are described below. Following these highlights, the specific objectives and progress in each work unit are reported.

Low temperature processing is becoming increasingly important, especially for large area liquid crystal displays which are on glass substrates. Si films become amorphous when deposited below 600!C, thus limiting the quality of low temperature films for thin film transistors (TFT) and poly gates for the MOSFETs required for the active matrix switches in these displays. Poly SiGe films have been deposited by LPCVD at 500°C with 5X lower sheet resistance than for poly Si films and less than half the sheet resistance for poly Si films deposited at 900°C. These SiGe films have produced TFTs on glass with improved characteristics because of the higher mobility of SiGe compared to amorphous Si. This work has led to joint projects with TI and Motorola to exploit these advantages of SiGe in liquid crystal displays.

Recently, there has been great interest and excitement over the observation of efficient photoluminescence from porous Si. One explanation of the phenomena has been quantum confinement of the micro-structure Si particles in the porous matrix--much like quantum dots in GaAs/AlGaAs. We have fabricated precise dimensionally controlled, regular Si columns by a combination of e-beam lithography, etching and oxidation. While the fabrication is a remarkable achievement, there is virtually no light emission if these structures compared to the porous Si structures. This suggests that surface phenomena and Si-H bonds play a more critical role in this emission process rather than quantum confinement. This work will continue to focus on understanding the mechanism in these structures, as we can fabricate reproducible nanostructures, the fundamental building blocks and then treat the surfaces in different ways to simulate the porous Si surface.

Bipolar transistors have traditionally been utilized for low noise applications where $1/f$ noise plays a role. While there has been great progress in realizing much higher frequency performance in heterojunction bipolar transistors (HBT), there has been little work to determine how well these devices will perform in low noise applications. We have demonstrated the role of surface recombination and D-X centers in the AlGaAs emitter as major sources of $1/f$ noise. Fabrication of devices with reduced Al content in the emitter and AlGaAs base surface passivation ledges reduce the $1/f$ noise of these devices by 20 fold compared to conventional HBTs.

Along more mathematical lines, a complete mathematical determination has now been made of problems for which sequential data compression can be optimally achieved. Sequential data compression consists of first approximating data using a few bits of information, then interactively improving the approximation as more and more information is supplied. An ideal system would provide description which is rate distortion optimal at each state. This work shows that such an ideal solution is possible if and only if the individual solutions of the rate distortion problems at each stage of consistent with the marginals of a Markov chain, a condition that is easily checked. This work implies in particular that tree-structured descriptions are optimal if and only if the rate distortion problem is successively refinable. Also, a counterexample shows that not all problems are successively refinable. The implications of this work in HDTV will be to enable an image to be compressed optimally at a black and white level, then at a more refined color level, and finally at full HDTV precision, each stage being optimal given the

4

information supplied. The multistage data compression algorithm is actually simpler than the single stage algorithm. Some of the achievements of this work are currently being written into the international data standard.

A useful result has been obtained in the recovery of information from distributed memories. Using techniques from Galois field theory and network flow theory, we have been able to show how to distribute information among several different disks with a minimum of redundancy so that no data is lost even if several disks crash. This theory applies directly to IBM's plan to use multiple cheap disks instead of a single expensive disk. The work also demonstrates how to reduce the time spent updating parity checks disks when data disks are continually being modified.

The technical knowledge developed under the JSEP contract is widely disseminated through sponsor reviews, presentations of papers at technical meetings, publications in the open literature, discussions with visitors to the laboratories, and publication of laboratory technical reports (Ph.D. dissertations).

Unit: 1

TITLE: Investigation of High Speed Devices

Principal Investigator: J. S. Harris, Jr.

Graduate Students: D. Costa and P. Cheng

Scientific Objectives

The objective of this work is to extend our understanding of the AlGaAs/GaAs heterojunction bipolar transistor (HBT) in two important areas: high frequency operation and low frequency noise. High frequency device optimization has been limited in the past by lack of a physically based model which could be derived from high frequency S parameters. device optimization based on equivalent-circuit parameters. These parameters are obtained through a novel, physically based, de-embedding technique. The low-frequency noise characteristics of HBTs have been problematic for ultra-low noise oscillators and electronic receivers. We have examined HBTs as a function of various device-structure parameters under different operating conditions. In contrast to many previous studies, the origins of various noise components are interpreted in terms of specific physical quantities, such as surface recombination velocity and deep donor DX center concentration, and can thus be modified through device design.

Summary of Research

Equivalent-circuit de-embedding process.

Further improvement in the high-frequency performance of HBTs requires very careful optimization of the epitaxial structure, processing, and device geometry. Equivalent-circuit models are a useful tool for understanding the speed potential of HBTs, particularly when the device must simultaneously satisfy a number of performance criteria. Accurate, physically-based device models serve as a bridge between the known design/process variables and the desired high-frequency response [Costa]. Since any model is only as accurate as the accuracy of its element values, the issue of model-parameter extraction is of paramount importance. In this study, a new technique for directly extracting the small-

signal equivalent circuit model of AlGaAs/GaAs HBTs was developed. This model is shown schematically in figure 1. This procedure can be summarized as follows:

1) measure the s-parameters of the transistor embedded in its parasitic environment

2) convert the s-parameters to y-parameters and subtract the parasitic shunt elements C_{pbe}, C_{pbc}, and C_{pce}

3) convert the new y-parameters to z-parameters and subtract the parasitic series elements L_{pb}, R_{pb}, $R_{b,con} \parallel C_{b,con}$, L_{pe}, R_{pe}, R_E, L_{pc}, R_{pc}, and R_c

4) convert the new z-parameters to y-parameters and subtract the parasitic shunt element $C_{jc,ext}$

5) convert the new y-parameters to z-parameters and subtract the parasitic series element R_{bb}

6) convert the new z-parameters to y-parameters that correspond to the intrinsic HBT y-parameters.

Applying simple circuit analysis, we can directly compute the intrinsic element values from the intrinsic HBT y-parameters.

Physical Measurement of Parasitic Elements

The first step in the de-embedding process is to minimize the number of elements which must be fit in Fig. 1. By careful design and incorporation of test structures with the devices, one can independently determine most of the parasitic elements in Fig. 1. The parasitics associated with the pads were determined by measuring a test pattern which consisted of only the pads. The parasitic device-connection impedances were determined by measuring a test pattern which consists of the pads, the device feeds, and a short replacing the transistor. The series emitter resistance was determined by means of the open-collector method [Getreu]. The extrinsic base-collector parasitics were determined from RF measurements of a HBT structure whose emitter layer and base layer underneath the emitter had been removed by etching.

Modeling Results

The equivalent-circuit model can also be used to calculate the unilateral power and common-emitter current gains of the device and predict f_t and f_{max}. Figure 2 compares the measured unilateral power and common-emitter current gains with the experimental values for a device we tested. As can be seen from Fig. 2, an excellent fit is obtained.
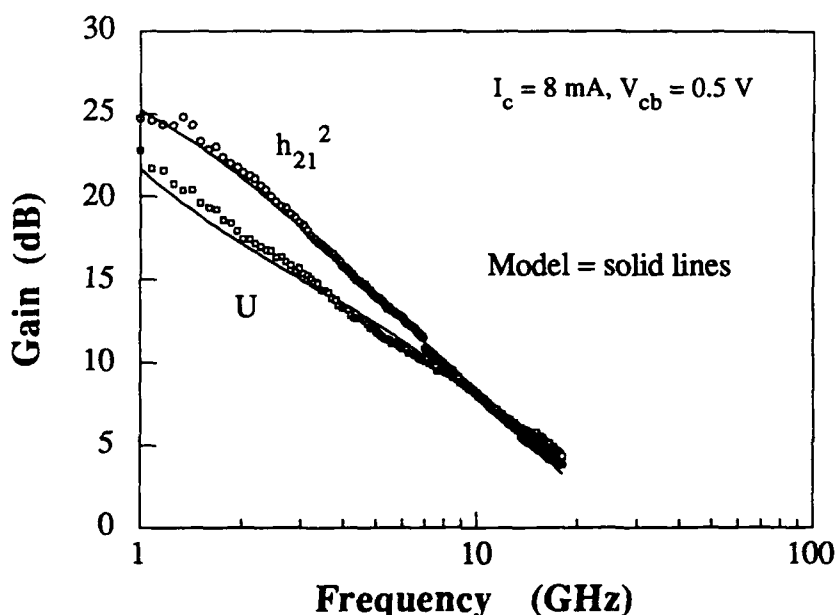
Figure 2   Comparison of the measured and model-produced unilateral power gain (U) and common-emitter short-circuit current gain (h_{21}) of a HBT biased at I_c = 8 mA and V_{cb} = 0.5 V.

The equivalent circuit obtained for the devices we studied provided many insights for device optimization. The most obvious information obtained was that the series emitter and collector resistances needed to be reduced. We also learned that the layout of the device and the device isolation are both quite suitable. The validity of this technique was confirmed by observing the frequency independence of the extracted circuit elements.

Characterization of Noise in HBTs

Low-frequency noise can limit the bandwidth and stability of a wide variety of integrated circuits [Chen]. It also degrades the spectral purity of nonlinear microwave circuits, such as

oscillators and mixers where the low-frequency, base-band noise up-converts as noise sidebands around the RF carrier signal [Siweris]. While the high-frequency performance of AlGaAs/GaAs heterojunction bipolar transistors (HBTs) is well-established [Lang], the noise properties have largely been overlooked. Low- frequency noise measurements (100 Hz to 10 MHz) of Npn AlGaAs/GaAs HBTs with and without AlGaAs surface passivation ledges were made as a function of bias current, device geometry, aluminum mole fraction in the emitter, and temperature. Figure 3 shows schematic cross sections of HBTs (a) without the AlGaAs passivation ledge and (b) with the AlGaAs passivation ledge. This ledge allowed us to measure the effect of base surface recombination on noise; without the ledge, the depletion region formed on the exposed GaAs base accelerates minority carriers towards the surface where they immediately recombine. The mask design was such that for a given device geometry, devices with and without the ledge were adjacent to each other.

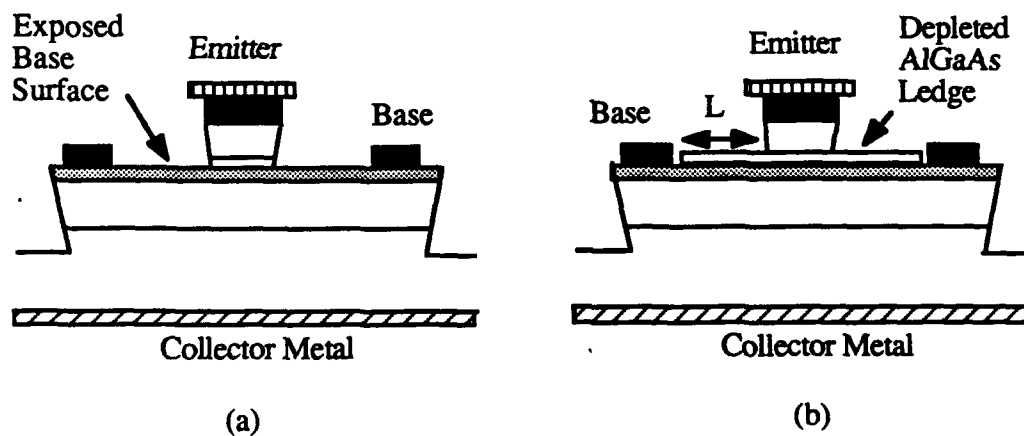
(a)   (b)

Figure 3   Schematic cross sections of HBTs (a) without and (b) with the AlGaAs surface passivation ledge covering the extrinsic base surface.

The measured noise spectra arise from a number of different sources, which can each be classified by their spectral content. White noise refers to noise whose spectral density is independent of frequency. The major source of white noise associated with the input base noise current is due to shot noise from the dc base current. Shot noise arises from the discrete nature of electronic charge and the fact that the current flow consists of individual pulses, which flow with random spacing. 1/f noise derives its name from the fact that its spectral density is roughly inversely proportional to frequency. Although many mechanisms (both bulk and surface related) for 1/f noise have been suggested, its physical origins have not been conclusively determined, even in conventional homojunction silicon bipolar transistors. Many theoretical studies have associated 1/f noise with surface states [Fonger; McWhorter; Jantsch; van der Ziel]. Numerous experimental results support this surface noise theory [Sah; Hsu; Khajezadeh; Chung, Radford; Hayama]. Electrical noise in the form of random "bursts", caused by a switching between two states (in the simplest case), is sometimes observed in semiconductor devices. Burst noise in bipolar transistors is often associated with traps or generation-recombination (g-r) centers near the emitter-base space charge region.

Typical noise measurements for our HBTs are shown in Fig. 4 and they show the existence of three distinct regions in the noise spectra: a 1/f line-shape at the lower end of the measured frequency range, a Lorentzian spectrum (noise "bump") at intermediate frequencies, and a white noise region at the higher end of the measured frequency range.

11
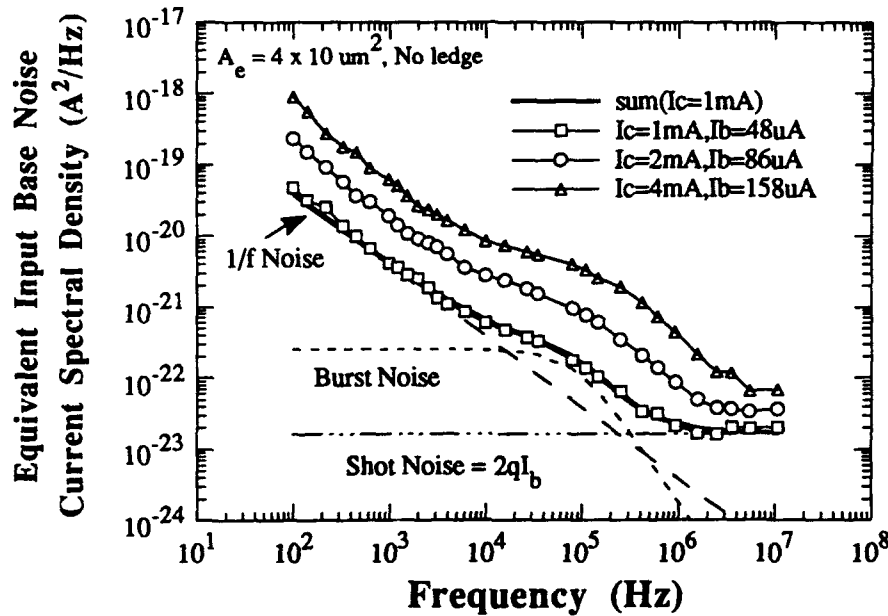
Figure 4. Equivalent input base noise current spectral densities and the various noise components for a 4 μm x 10 μm emitter $Al_{0.3}Ga_{0.7}As/GaAs$ HBT without the AlGaAs passivation ledge and operating in the linear region at three different bias currents.

2.2 Noise Results

As is well-known, AlGaAs has a number of electron traps because Al is more reactive than Ga [Yamanaka]. Because burst noise in bipolar transistors is usually attributed to defects in the emitter-base space charge region and the depletion region mainly extends into the more lightly doped AlGaAs emitter, we measured the low-frequency noise of HBTs with Al mole fractions of x = 0.2 and x = 0.3 in the emitters. Figure 5 compares the equivalent input base noise current spectral densities of 4 μm x 10 μm emitter $Al_{0.3}Ga_{0.7}As/GaAs$ HBTs with those of similar size $Al_{0.2}Ga_{0.8}As/GaAs$ HBTs at a collector current density of 5 x 10^3 A/cm^2.
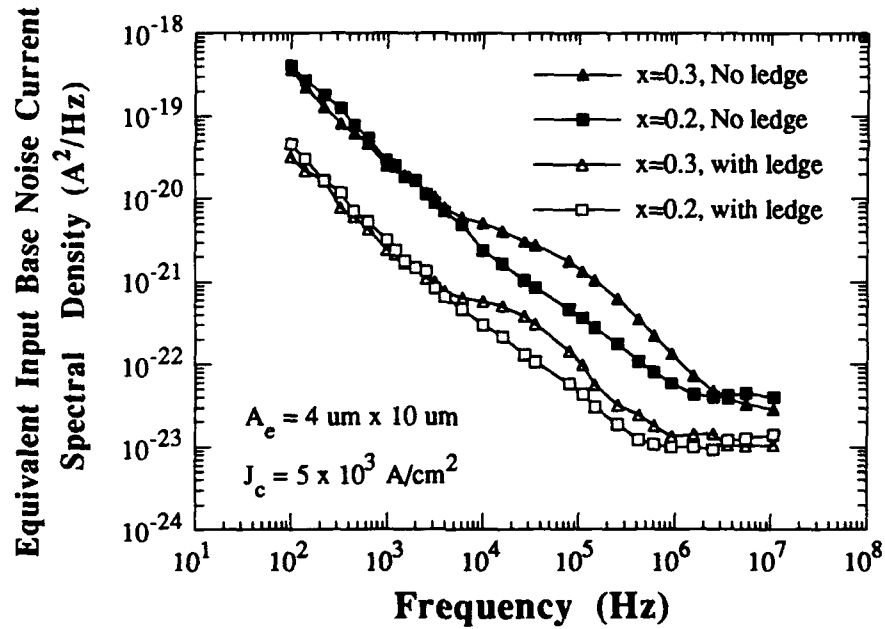
Figure 5 Comparison of the equivalent input base noise current spectral densities of 4 μm x 10 μm emitter HBTs with and without the AlGaAs passivation ledge for two different Al mole fractions x = 0.2 and x = 0.3, at a fixed collector current density of 5 x 10^3 A/cm^2.

Both devices with and without the AlGaAs ledge were compared. While the x = 0.2 and x = 0.3 devices exhibited comparable 1/f noise and white noise levels, the noise "bump" was virtually eliminated for the x = 0.2 devices. For both devices with and without the AlGaAs ledge, S_{I_b} (50 KHz) of the x = 0.2 devices was roughly three times lower than that of the x = 0.3 devices. These results are very strong evidence that the noise "bump" is associated with traps in the AlGaAs emitter.

The dominant source of 1/f noise in our AlGaAs/GaAs HBTs was shown to be associated with states at the extrinsic base surface. This finding is consistent with Fonger's [Fonger] model which attributes the 1/f noise to fluctuations in the surface recombination velocity and hence fluctuations in the extrinsic base surface recombination current. Our measurements also suggest that the origin of the anomalous noise "bump" observed at intermediate frequencies is due to fluctuations in the occupancies of AlGaAs traps near the surface of the emitter-base

junction. The DX center was identified as a possible candidate for this trap. The results of this study suggest several modifications in the epitaxial structures and the fabrication processes that substantially reduce the low-frequency noise of AlGaAs/GaAs HBTs. We have demonstrated that the use of a protective AlGaAs ledge over the extrinsic base surface and a reduction of the Al mole fraction in the emitter from 0.3 to 0.2 significantly improves the low-frequency-noise performance of our AlGaAs/GaAs HBTs. Based on these observations, there are several structures and materials which could further improve the low-frequency noise performance of HBTs, such as:

- HBTs utilizing n-type emitters composed of materials known to be free of DX centers, or Pnp AlGaAs/GaAs HBTs, since DX centers do not exist in p-type AlGaAs.

- Inverted collector-up/emitter down HBTs where the buried base-emitter junction eliminates the exposed surface suspected of being the noise source.

- HBTs incorporating a thinner base and/or a quasi-electric built-in field in the base which reduce minority carrier diffusion to the exposed surface.

- Utilize HBT base layers composed of materials with lower surface recombination velocities than GaAs, such as InGaAs.

References

[Chen]   C. Y. Chen, J. Bayruns, R. J. Bayruns, and N. Scheinberg, "Reduction of Low-frequency Noise in a DC-2.5 GHz GaAs Amplifier," *GaAs IC Symposium Digest*, p. 289, 1988.

[Chung]   H. K. Chung, M. A. Rosenberg, and P. H. Zimmerman, "Origin of 1/f nosie in HgCdTe Photodiodes," J. *Vac. Sci. Tech. A*, vol. 3, No. 1, p. 189, 1985.

[Costa]   D. Costa, W. Liu, and J. S. Harris, Jr., "Direct Extraction of the AlGaAs/GaAs Heterojunction Bipolar Transistor Small-Signal Equivalent-Circuit," *IEEE Trans. Electron Dev.*, Sept. 1991.

[Fonger]   W. Fonger, " A determination of 1/f noise sources in semiconductor diodes and transistors," in *Transistors I*, Princeton, NJ: RCA Labs, p. 239, 1956.

[Getreu]   I. E. Getreu, *Modeling the Bipolar Transistor*, pp. 140-150, 1978.

[Hayama]   N. Hayama, S. Tanaka, and K. Honjo, "1/f Noise Reduction for Microwave Self-aligned AlGaAs/GaAs HBTs with AlGaAs Surface Passivation Layer," *Third Asia-Pacific Microwave Conference Proceedings*, p. 1039, 1990.

[Hsu]   S. T. Hsu, "Surface State Related 1/f Noise in p-n Junctions," *Solid State Elect*, vol. 13, p. 843, 1970.

[Jantsch]   O. Jantsch, "Flicker (1/f) Noise Generated by a Random Walk of Electrons in Interfaces,"*IEEE Trans. Electron Devices*, ED-34, p. 1100, 1987.

[Khajezadeh] H. Khajezadeh and T. T. McCaffery, "Materials and Process Considerations for Monolithic Low-1/f-Noise Transistors," *Proc. IEEE*, vol.57, No.9, p.1518, 1969.

[Lang]   D. V. Lang, R. A. Logan, and M. Jaros, "Trapping characteristics and a donor-complex (DX) model for the persistent-photoconductivity trapping center in Te-doped AlxGa1-xAs," *Phys. Rev. B*, vol. 19, No. 2, p. 1015, 1979.

[McWhorter] A. L. McWhorter, *Semiconductor Surface Physics*, Philadelphia, p.169, 1957.

[Radford]   W. A. Radford and C. E. Jones, "1/f noise in Ion-Implanted and Double-Layer Epitaxial HgCdTe Photodiodes," *J. Vac. Sci. Tech. A*, vol. 3, No. 1, p. 183, 1985.

[Sah]   C. T. Sah and F. M. Heilscher, "Evidence of Surface Origin of the 1/f Noise," *Phys. Rev. Lett*, vol. 17, p. 956, 1966.

[Siweris]   H. J. Siweris and B. Schiek, "Analysis of Noise Upconversion in Microwave FET oscillators," *IEEE Trans. Microwave Theory Tech.*, vol. 33, p.233, 1985.

[van der Ziel]   A. van der Ziel, "Formulation of Surface 1/f Noise Processes in Bipolar Junction Transistors and in p-n Diodes in Hooge-Type Form," *Solid State Elect*, vol. 32, No. 1, p. 91, 1989.

[Yamanaka]   K. Yamanaka, S. Naritsuka, K. Kanamoto, M. Mihara, and M. Ishii, "Electron traps in AlGaAs grown by molecular-beam-epitaxy," *J. Appl. Phys.*, vol. 61, No. 11, p. 5062, 1987.

JSEP Supported Publications

1. S. J. B. Yoo, M. M. Fejer, R. L. Byer and J. S. Harris, Jr., "Second Susceptibility in Asymmetric Quantum Wells and its Control by Proton Bombardment," *Appl. Phys. Lett.* **58**, pp. 1724-1726, April 1991.

2. D. Costa and J. S. Harris, Jr., "Low-Frequency Noise Characterization of Npn AlGaAs/GaAs Heterojunction Bipolar Transistors", *Proceedings of the 18th International Symposium on GaAs and Related Compounds*, September 1991, Seattle, WA.

JSEP Supported Thesis

P. Cheng, "Novel Tunneling Barrier Designs for Resonant Tunneling Diodes," Ph.D. Thesis, Stanford University, January 1991.

D. Costa, "Microwave and Low-Frequency Noise Characterization of Npn AlGaAs/GaAs Heterojunction Bipolar Transistors," Ph.D. Thesis, Stanford University, December 1991.

Unit: 2

TITLE: Physics and Applications of Ultra-Small Materials Structures

SENIOR INVESTIGATOR: R. F. W. Pease

RESEARCH STUDENTS: H. Liu and T. Takken

Scientific Objectives

The overall objectives of this program are to investigate opportunities for new devices whose operation depends on quantum mechanical effects associated with the ultrasmall (<100nm) nature of these devices. Obviously we hope to realize device concepts that promise superior performance than that achieved by devices based on current principles. Even if that desirable end is not achieved we should in any case be in a position to optimize the design of devices employing lateral features below 100 nm when our ability to fabricate them moves beyond the laboratory into manufacturing.

Summary of Research
Ultrasmall Semiconductor Structures

The primary effort during the current reporting period has been the investigation of the photoluminescence of ultrasmall silicon structures. This work has been performed in cooperation with members of the staff of Xerox, Palo Alto Research Center.

The reported photoluminescence of porous (etched) silicon is significant; both because of the possibility of realizing optical sources in silicon, and because of the appealing explanation that the microscopic nature of the silicon between the pores in per se responsible for the effect. One thesis is that the quantum confinement of the carriers in such microstructure led to the possibility of radiative transitions. However, less appealing explanations also exist; notably those that suggest chemical effects, such as hydrogenation as a result of the etching.

We have reviewed the published work, demonstrated clearly the photoluminescence of porous silicon ourselves and performed selected experiments to help in finding an explanation. A "scorecard" has been drawn up of evidence for and against the quantum-confinement explanation.

17

In addition we have generated arrays of silicon nanostructures by means (electron beam lithography, reactive ion etching and oxidation) other than porous etching to see if these well-controlled structures also luminesce. Quite apart from the fabrication difficulty, the photoluminescence experiment is also not straightforward in that we initially obtained a false positive response which turned out be due to a diffracted beam from the regular array of nanostructures which caused the collection lens to luminesce.

This series of experiments is not only aimed at finding out whether the quantum-confinement in porous silicon is the responsible mechanism but under what circumstances lithographically defined silicon can be made to luminesce.

Score Card

a) Supporting the quantum confinement hypothesis

1. TEM images of porous Si show the existence of sub-10 nm microcrystallites. The diffraction patterns from those crystallites are similar to those of a single crystalline Si wafer.

2. The photoluminescence spectra of porous Si exhibits a blue-shift the longer the sample etched, supporting the theory that smaller structures give a larger split in subband energies. Similar blue-shifts in spectral peaks were observed when one increased the measurement temperature, which is a common phenomenon in a 2-D quantum well system with imperfect confining interfaces [Gardelis]. Moreover, the PL intensity was shown to be have a linear dependence on the input laser power, consistent with the theory that the optical transition is the result of excitonic decay [Gardelis].

3. The technique of using oxidation to further shrink the Si nanostructures has been used on porous Si to show the blue-shift of PL spectra [Shih]. In addition to the blue-shift, Shih *et al.* also observed the narrowing of the spectral width after oxidation, which seems to suggest that oxidation can be used to reduce the size distribution of Si nanostructures.

4. Annealing porous Si in Ar at 800 °C for 1 hour showed that the PL intensity not only did not attenuate but became stronger than that of the unannealed samples. This temperature should be sufficient

18

to drive out most impurities, suggesting that the luminescence is probably coming from the Si itself.

5. Room temperature visible electro-luminescence of porous Si has also been demonstrated [Koshida]. A thin semitransparent metal (Au) layer was evaporated on top of porous Si to establish contact and to pass current through this simple diode structure. The device exhibits sub-exponential rectifying current voltage characteristics and luminesces when the forward current exceeds a critical value. The luminescence spectra are similar to that obtained from PL. Moreover, the luminescence intensity increases nearly in proportion to the forward current density, suggesting that photon emission is based on a carrier injection mechanism.

6. Si microcrystals prepared by sputtering in a hydrogen plasma [Furukawa] and Ge microcrystals prepared by annealing Ge doped glass (SiO_2) [Maeda] also exhibit room temperature visible photoluminescence. Both authors interpret their results as quantum confinement effects.

7. Visible light emission has also been observed from lithographically defined and RIE etched Ge pyramid structures with sharp edges [Venkatasubramanian]. However, in our opinion, this result is somewhat questionable. We had similar experience with sharp cone structures in Si, but it turns out to be a bizarre case of artifact. The luminescence was coming from the filter put in front of the monochrometer to eliminate laser plasma lines, not from the Si structures.

b) Against the quantum confinement hypothesis

1. Besides the scattered Si microcrystallites, TEM images of porous Si also show a huge amorphous background. It is well known that efficient visible luminescence with PL spectra similar to that of porous Si can also be obtained from hydrogenated amorphous Si [Wolford]. Moreover, the blue-shift of the PL spectra can also be accomplished by incorporating a greater percentage of hydrogen into amorphous Si.

2. Another chemical compound, siloxene ($Si_6O_3H_6$), has recently been demonstrated to show PL and IR absorbence spectra similar to that of porous Si [Brandt].

3. Si-rich SiO_2 can also be made to luminesce efficiently at room temperature [DiMaria]. It is possible that such a compound is abundant in porous Si when it is exposed to air, because of its large reactive surface area.

4. Luminescence degradation in porous Si during PL measurement has been observed by us and other workers [Tischler]. A more detailed ambient study by Tischler *et al.* showed that illumination in the presence of O_2 introduces a significant number of dangling bond defects and that the presence of passivating hydrogen atoms is essential for obtaining high luminescence efficiency. From this experiment, it seems that the PL from porous Si can be explained by surface Si-H alloying effects.

5. From the spectral width of a porous Si PL spectrum, one finds that if one were to use the simple particle in a box model, one would estimate a size distribution of less than 1 nm. For a network of pores as complicated as that exists in porous Si, this seems difficult to believe, unless there is some self-limiting process in anodic etching.

Discussion

From the above, we see that the exact mechanism responsible for the visible luminescence from porous Si is far from clear. In fact, we believe that it will be very difficult to obtain definite answers from working with porous Si alone. Ideally, one needs to be able to start with more regular artificially defined Si nanostructures and then introduce each element that may be responsible for the luminescence independently. To get some idea of the structural size required to observe the quantum effect, one can estimate the Bohr radius of excitons in silicon, $a_B = \varepsilon h^2 / \pi \mu e^2$. For bulk Si, this is on the order of 30 Å, which means that one needs to be able to achieve structures comparable to this dimension to observe effects due to confined excitons.

Toward this end, we have fabricated Si nanostructures to emulate the proposed quantum structures in porous Si. The process involved using a custom made high resolution electron beam lithography system in combination with a lift-off process to first define nanometer scale Cr dots on top of a thermally grown oxide layer. The Cr dot patterns were then transferred onto the oxide layer by CHF_3 RIE. Finally, the oxide mask was used to etch Si by RIE with Cl_2 and $SiCl_4$. This process allows us to make Si columns as

small as 25 nm in diameter and as 0.25 μm high. An example of this is shown in Fig. 1.

We then use oxidation to further reduce the (silicon) structure size. We have also demonstrated that it is possible to monitor the oxidation progress by taking advantage of the diffraction contrast between Si and SiO_2 and using a TEM to image the Si core surrounded by an oxide skin (Fig. 2). At this time, although we believe that we have achieved Si core size as small as 50 Å, we have not seen any PL signals from these structures at room temperature.

From the evidence gathered on porous Si, our sense is that size is not the only issue. It seems that surface passivation to eliminate nonradiative recombination channels could also be very important. Therefore, we are now working on a combination of annealing and hydrogenation to passivate the damages due to the RIE processes. Moreover, we are also working on the photoemission facility at SLAC to study the band structure of both porous Si and the artificial Si nanostructures. Making electrical contacts to the lithographically defined Si nanostrucutres is another interesting experiment that we are pursuing at this time. Hopefully, these studies will yield more information about the physics behind the luminescence of those tiny microcrystallites and contribute to the understanding of the electronic structure of nanostructures made of indirect bandgap materials.
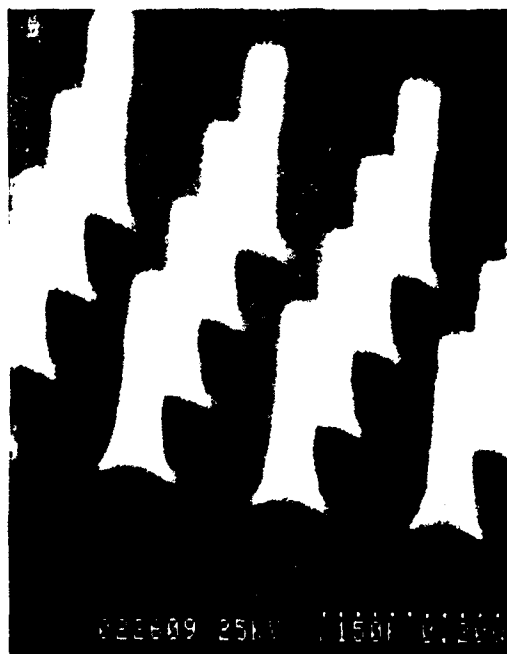

Figure 1   SEM micrograph of Si nanostructures
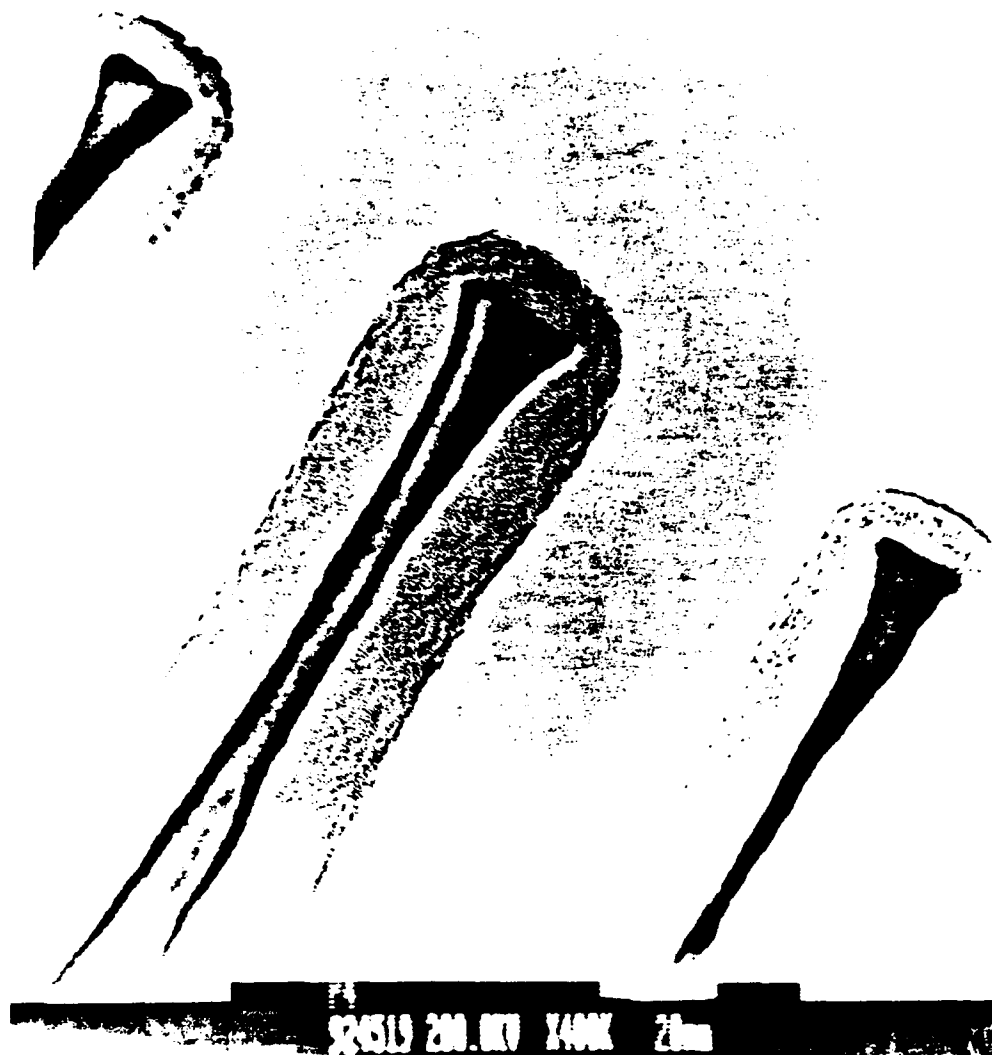
Figure 2   TEM micrograph of oxidized Si nanostructures

Ultra Small Superconductor Structures

a) Weak link formation in high temperature superconductors using high resolution patterning.

This project has been entirely supported during the current reporting period by contract F49620-90-C-0081 (AFOSR).

b)   Ultra Small Superconducting Transmission Lines

This project is just being initiated in cooperation with Prof. Beasley. In previous work, the dimensions of the lines have been in excess of the London penetration depth and with values of current density much less than the critical value.

We are interested in the behavior of devices, circuits and systems with dimensions below 100nm. These dimensions will be of the order of, and in some cases even smaller than, the reported values of the London penetration depth. In addition, these small dimensions will lead to current densities that may well exceed the critical values.

22

As an initial project we are building on an earlier project on superconducting transmission lines in the size range $\geq 1\mu m$ by
1) shrinking the dimensions to 0.1μm with conventional materials
2) Changing the material to an alloy that exhibits very large penetration depth (~10μm); in this way, we can identify any special characteristics (e.g. excessive attenuation, non-linearly) associated with dimensions being less than the penetration depth. In each case the lines will be characterized using the resonator technique used in the earlier work [Langley].

References

[Brandt]   M. S. Brandt, H. D. Fuchs, M. Stutzmann, J. Weber and M. Cardona, preprint submitted to *Solid State Communications* (1991).

[Canham]   L. T. Canham, *Appl. Phys. Lett.* **57** 1046 (1990).

[DiMaria]   D. J. DiMaria, J. R. Kirtley, E. J. Pakulis, D. W. Dong, T. S. Kuan, F. L. Pesavento, T. N. Theis, J. A. Cutro, and S. D. Brorson, *J. Appl. Phys.* **56** 401 (1984).

[Furukawa]   Shoji Furukawa and Tatsuro Miyasato, *Japanese Journal of Applied Physics* **27** L2207 (1988).

[Gardelis]   S. Gardelis, J. S. Rimmer, P. Dawson, and B. Hamilton *Appl. Phys. Lett.* **59** 2118 (1991).

[Koshida]   Nobuyoshi Koshida and Hideki Koyama, *Appl. Phys. Lett.* **60** 347 (1992).

[Langley)   B. W. Langley, S. M. Anlage, R. F. W. Pease, and M. R. Beasley, *Rev. Sci. Instrum.* **62** (7), (July 1991).

[Maeda]   Yoshihito Maeda, Nobuo Tsukamoto, and Yoshiaki Yazawa, *Appl. Phys. Lett.* **59** 3168 (1991).

[Shih]   S. Shih, C. Tsai, K.-H. Li, K. H. Jung, J. C. Campbell, and D. L. Kwong, *Appl. Phys. Lett.* **60** 633 (1992).

[Tischler]   M. A. Tischler, R. T. Collins, J. H. Stathis, and J. C. Tsang, *Appl. Phys. Lett.* **60** 639 (1992).

[Venkatasubramanian] R. Venkatasubramanian, D. P. Malta, M. L. Timmons, and J. A. Hutchby, *Appl. Phys. Lett.* **59** 1603 (1991).

[Wolford] D. J. Wolford, B. A. Scott, J. A. Reimer and J. A. Bradley, *Physica* 117B 920 (1983).

JSEP Supported Publication

JSEP Supported Thesis

Unit: 3

TITLE: Reactive Ion Profiling of Heterostructures

PRINCIPAL INVESTIGATOR: C. R. Helms

GRADUATE STUDENTS: M. Kniffin,
T. Beerling and R. Keller

Scientific Objectives

The objective of this work was to determine the surface chemistry associated with the interaction of reactive ions with GaAs and other III-V surfaces. In-situ electron spectroscopy and mass spectroscopy techniques developed for similar studies of SiO_2 and Si [Thomson '85], [Thomson, '90], [Thomson '86] were employed in this investigation.

Summary of Research

Since the program began, new results have appeared indicating the utility of low temperature deposition and etching in downstream plasma reactors (in both RF and ECR microwave configurations). With this in mind, we have constructed a UHV compatible downstream plasma source which was intended for the III-VI Surface Studies. It became apparent, however, that such a system would be useful for studies related to diamond nucleation and growth, and our results to date on this work will be presented indicating an important new research direction.

Downstream Plasma Processing

Past work using downstream plasma processing for low temperature depositions have resulted in operating MIS devices on GaAs [Fountain] as well as high quality Si MOS devices [Richard] using low temperature grown oxides. These existing results have led us to change direction in this program to construct such a reactor on an existing UHV surface analysis system. In our first experiments with this approach the utility of such a reactor, operating with either hydrogen or oxygen, for prediamond growth surface processing will be assessed.

To perform these studies, a variety of in situ spectroscopic tools will be employed in a UHV analysis system. This study will focus on the following subjects; cleaning of Si and GaAs surfaces using hydrogen plasmas, passivation of GaAs surfaces, and pre-diamond growth surface treatments.
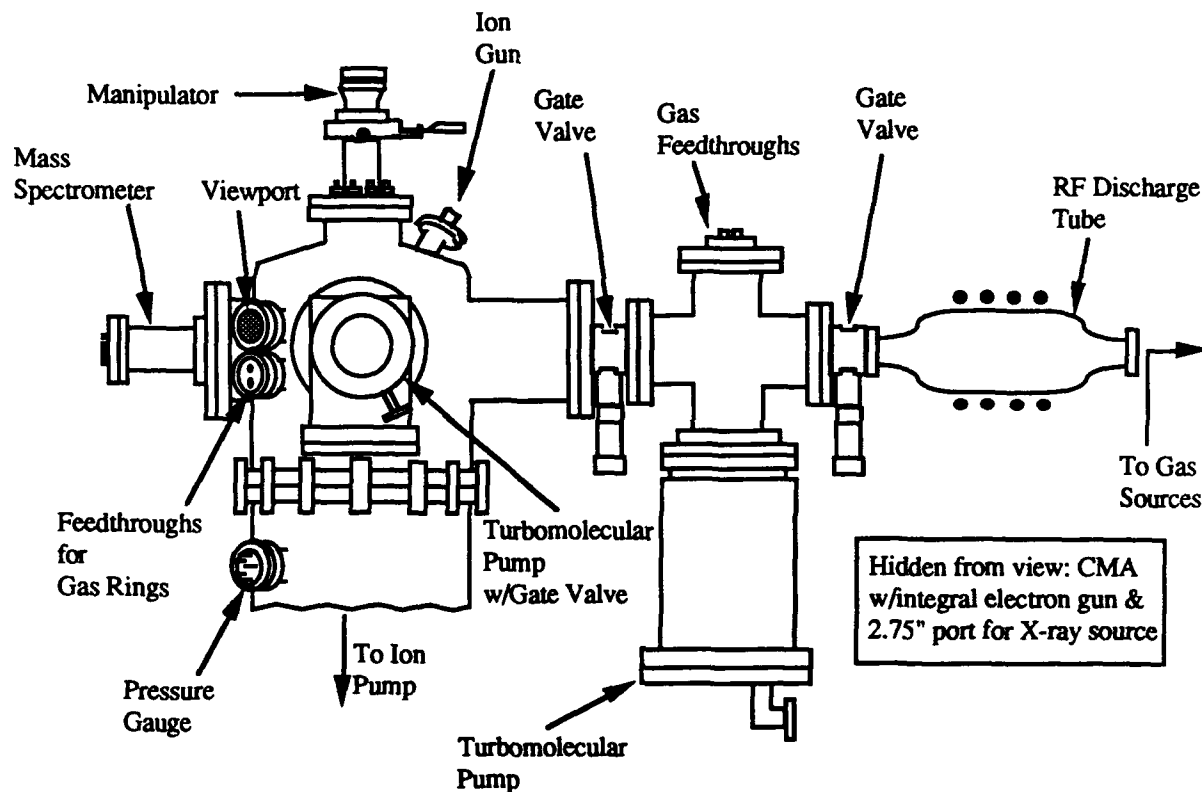
Figure 1. UHV system to be used in study.

Figure 1 is a diagram of the UHV system to be used in this study. Plasmas can be generated upstream from the samples either by an inductive coil, operating at 13.56 MHz, or by an Electron Cyclotron Resonance Source (ECR), operating at 2.45 GHz. At the present time, the system is equipped with an inductive coil. Since the pressures necessary to sustain plasmas are generally much higher than the maximum allowable pressures when operating many spectroscopic tools, a differential pumping stage, pumped with a turbomolecular pump, is used to reduce pressures in the main chamber to the required values. Pressures in the differential pumping stage and chamber are sufficiently low that excited species do not relax before reaching the samples.

In the chamber, a quadrapole mass spectrometer (QMS) is positioned line-of-site with the plasma so the excited species in the gas phase may be analyzed. With the heater assembly on the sample tray, the QMS can also be used to perform Temperature Programmed Desorption studies (TPD). To analyze the samples, Auger

Electron Spectroscopy (AES) or X-Ray Photoemission Spectroscopy (XPS) may be performed. The chamber is presently equipped with an AES system, although a double pass CMA and X-ray source are readily available.

As a part of this work, we have investigated carbon standards (graphite, natural diamond, synthetic diamond) with Auger Electron Spectroscopy (AES) and Electron Energy Loss Spectroscopy (EELS) and begun a study of the effects of annealing and ion damage on surface structure. Significant results include the finding that AES and EELS used in combination, provide a powerful method of characterizing the bonding nature of the first few nanometers of diamond and related surfaces.

The utility of using AES for characterizing diamond-like surfaces is illustrated in Fig. 2, where spectra of natural diamond, synthetic diamond (obtained from Crystallume, Inc.), diamond after sputtering and high temperature annealing, and graphite are shown. The natural diamond spectra obtained after boiling in concentrated H_2SO_4 is similar to that reported for diamond (111) 1 x 1 surfaces which are known to be hydrogenated. The synthetic diamond surface appears similar to dehydrogenated surfaces which exhibit more complex surface structures [Hamza, Kubiak 88, Kubiak 89, Pate]. Surface damage (via Ar^+ sputtering) and *in situ* high temperature annealing produces spectra with obvious graphite like features indicating that the surface region now contains significant π bonding [Williams]. First derivative Electron Energy Loss (EELS) spectra corresponding to the AES spectra of Fig. 2 are shown in Fig. 3. Both the natural and synthetic diamond spectra show features at 38 and 25 eV (previously reported at 35 and 23 eV in N(E) data) [Belton, Wang]. The annealed, damaged, surface shows a weakening of these features, a general broadening and the appearance of the strong 8 eV feature (6 eV in N(E)) associated with graphitic π bonding. The 30 eV graphitic feature is not well resolved due to interference from the diamond-like components remaining in the spectra. Future work investigating the interaction of atomic hydrogen with such surfaces and the effect on ion damage are underway.

Future Work - Surface Passivation of GaAs

Studies of surface passivation of GaAs surfaces, using hydrogen plasmas, have been performed. [Gottscho] *et al* have shown increases in GaAs photoluminescence (PL) intensities during exposure to a hydrogen plasma. This increase in PL intensity implies a reduction in surface recombination velocity, which in turns suggests a reduction in the interface state density. Atomic
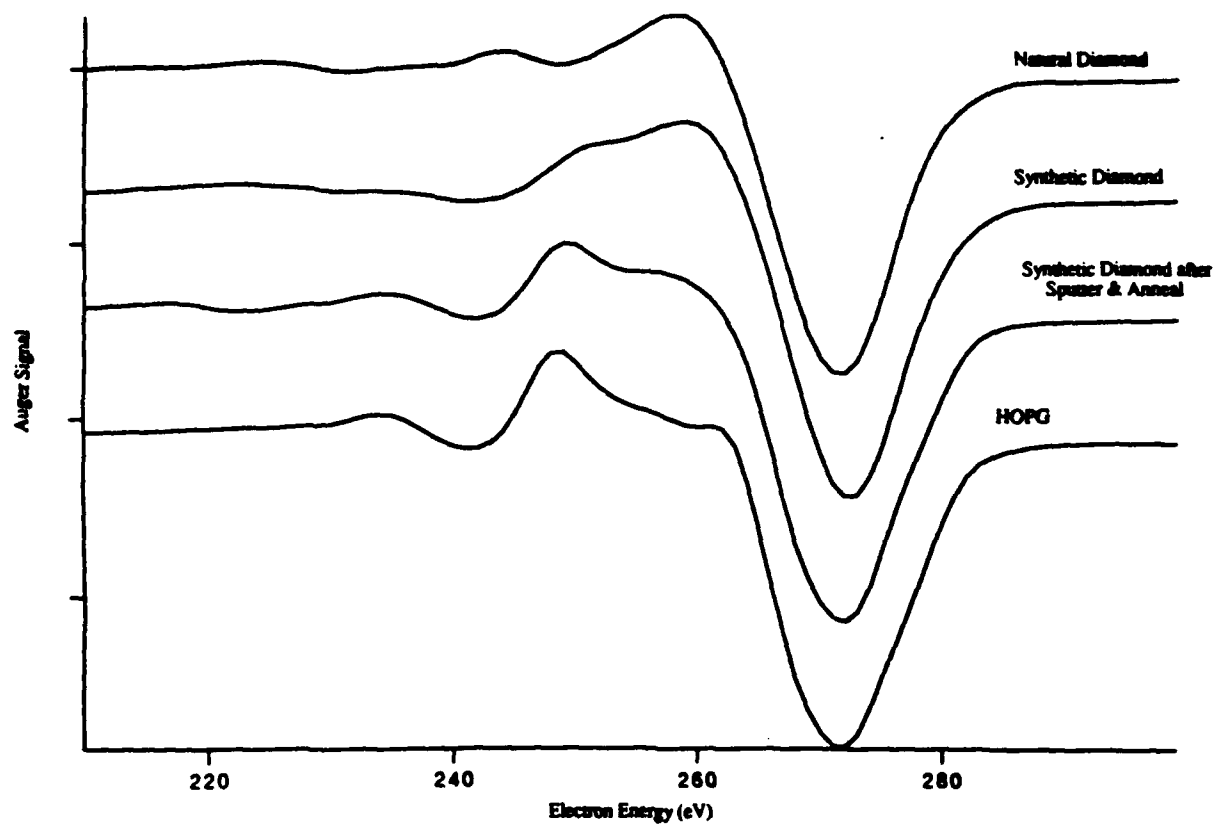
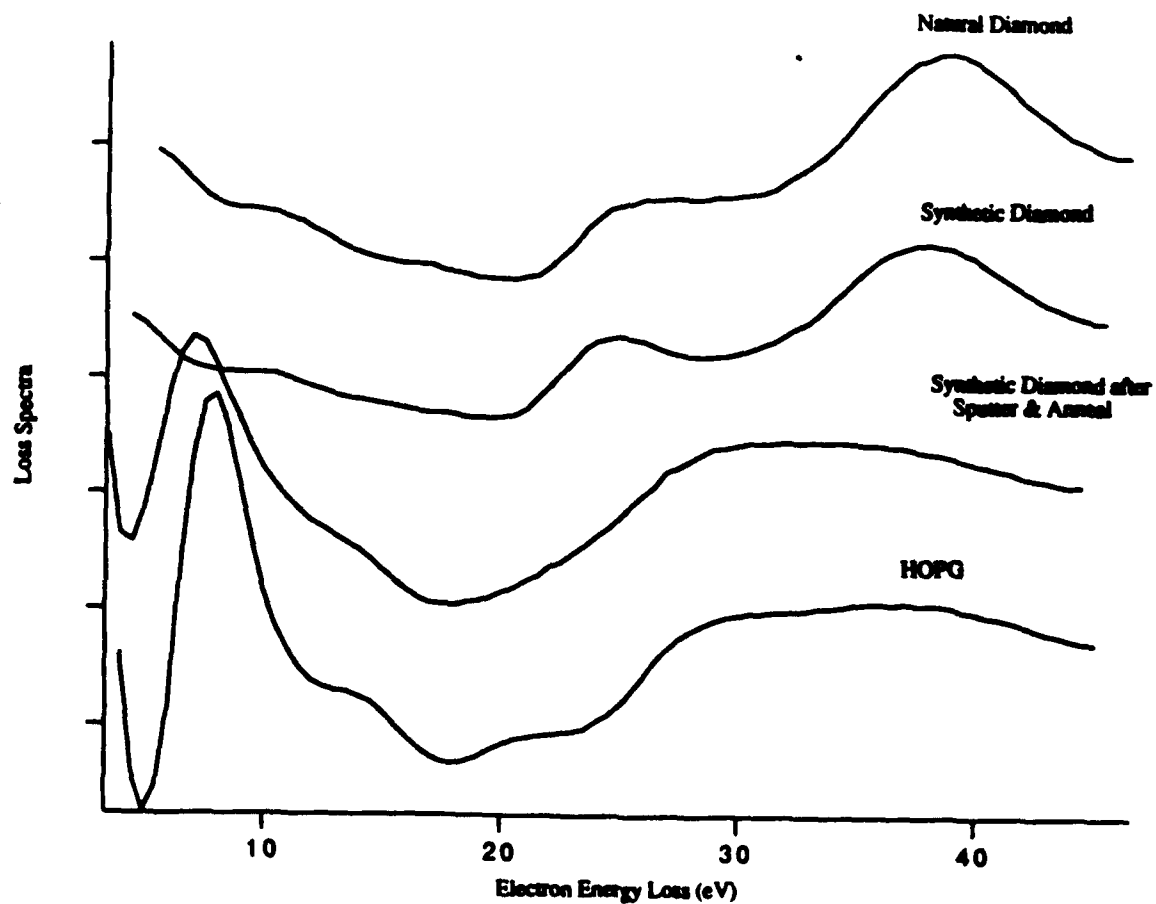Figure 2   Auger spectra of diamond, diamond-like films and graphite.

Figure 3   Loss Spectra of diamond, diamond-like films and graphite.

hydrogen, supplied by the plasma, is removing excess As (in the form of AsH3) from the the GaAs surface [Capasso]. This excess As at the GaAs surface is believed to be the cause for fermi level pinning at the GaAs surface [Callegari]. GaAs MOS structures have been built [Callegari] where GaAs surfaces are first exposed to a hydrogen plasma to reduce the As surface concentration and then exposed to a nitrogen plasma to grow a thin nitride layer. Gallium oxide was then deposited on the nitride layer. Interface state densities for these structures are believed to be about 10^{11} eV^{-1} cm^{-2} (an exact value could not be determined as the oxide was very leaky).

In our work, we wish to further pursue the subject of GaAs surface passivation. Work is to be performed where, along with hydrogen and nitrogen plasma treatments, other gases, such as ammonia (NH3), will be used. XPS will be used to determine the chemical effects of such a treatment. Structures to further examine the effects of the plasma treatment will be constructed. Metals of differing work functions will be deposited onto GaAs samples that have gone through a processing sequence which both removes excess arsenic and creates an ultrathin nitride layer, thin enough to permit tunneling. Barrier heights will be determined to see if the fermi level has been unpinned. MOS structures will be fabricated, but with oxides superior to the gallium oxide used in the work of [Callegari] so that a more reliable value for interface state densities may be determined.

References

[Belton]   D. N. Belton, S. J. Schmieg, *J. Vac. Sci. Tech.* **A8**, 2353, (1990).

[Capasso]   F. Cappaso, G.F. Williams, *J. Electrochem. Soc.*, **129** (4), 821, (1982).

[Callegari]   A. Callegari, P. D. Hoh, D. A. Buchanan, D. Lacey, *Appl. Phys. Lett.*, **54** (4), 332, (1989).

[Fountain]   G. G. Fountain, R. A. Rudder, S. V. Hattangady, R. J. Markunas, P.S. Lindorme, *J. Appl. Phys.*, **63** (9), 4744, (1988).

[Gottscho]   R. A. Gottscho, B. L. Preppernau, S. J. Pearton, A. B. Emerson, K. P. Giapis, *J. Appl. Phys.*, **68** (2), 440, (1990).

[Hamza]   A. V. Hamza, G. D. Kubiak, R. H. Stulen, *Surf. Sci.* **206** L833, (1988).

[Kubiak,88]   G. D. Kubiak, K. W. Kolanski, *J. Vac. Sci. Tech.* **A6,** 814 (1988).

[Kubiak,89]   G. D. Kubiak, K. W. Kolansinski, *Phys Rev.* **B39,** 1381 (1989).

[Pate]   B. B. Pate, *Surface Science* **165,** 83 (1986).

[Richard]   P. D. Richard, R. J. Markunas, G. Lucovsky, G. G. Fountain, A. N. Mansour, D. S. V. Tsu, *J. Vac. Sci. Technol.* **A3,** 867 (1985).

[Thomson, 85] D. J. Thomson and C. R. Helms, *Appl.. Phys. Lett.* **46**(11), (1985).

[Thomson, 87] D. J. Thomson, Ph.D. Thesis (1987).

[Thomson, 90] D. J. Thomson and C. R. Helms, *Surf. Sci.* **236,** 41, (1990).

[Williams]   B. E. Williams, J. T. Glass, *J. Mater. Res.* **4,** 373 (1989).

[Wang]   Y. Wang, H. Chen, R. 'N. Hoffman, J. C. Angus, *J. Mater. Res.* **5,** 2378 (1990)

JSEP Supported Publications

1. E. Weiss, R. C. Keller and C. R. Helms, "Formation of Single- phase MAs_x Films on GaAs by Selective Wet-hydrogen Oxidation and Etching"*J. Appl. Phys.* **69** (4), 2623 (1991).

2. R. C. Keller and C. R. Helms, "Calculations of the Barrier Height and charge Distribution of a Metal-Dielectric Interface", *J. Vac. Sci. & Technol.,* in press.

JSEP Supported Presentations

R. C. Keller and C. R. Helms, "Calculations of the Barrier Height and charge Distribution of a Metal-Dielectric Interface", 1991 AVS Meeting, Seattle, Washington.

Unit: 4

TITLE: The Electronic Structure and Interfacial Properties of High Temperature Superconductors

PRINCIPAL INVESTIGATOR: Prof. W. E. Spicer

JUNIOR FACULTY: Prof. Z.-X. Shen

GRADUATE STUDENTS: D. S. Dessau, B. O. Wells, D. M. King, D. S. Marshall, E. R. Ratner, J. DiCarlo

Scientific Objective:

The recently discovered copper-oxide based high temperature superconductors are of great interest both from a scientific and a technological viewpoint. The objective of this work is two-fold: first, to provide a better understanding of the electronic structure of these and related materials, ultimately aimed at gaining an understanding of the interfaces of the high temperature superconductors with other technologically important materials.

Summary of Research:

Our research has focussed on the applications of photoemission spectroscopy to the study of the electronic structure and interfacial properties of high temperature superconductors. In a way this research effort has been two-pronged; there has been a fundamental side to the research, with the ultimate goal being the elucidation of the mechanism of high temperature superconductivity, and there has been a technology side that is applications oriented.

The primary technique that we have applied to the study of the high temperature superconductors is angle-resolved photoemission spectroscopy (ARPES), which is the most direct method of determining the electronic structure of materials. In a photoemission experiment high energy photons (10-1000 eV) and absorbed and eject electrons from a sample surface, and these so-called photo-electrons are collected and then energy and momentum analyzed. To first order, the electron-removal spectrum tells us the one-electron density of occupied states of the system. Due to the short (5-10 Å) escape depth of the photoelectrons emitted from the sample surface, photoemission is a very surface sensitive tool. This surface sensitivity can be both an asset and a liability.

Recently we have been studying the high temperature super-conductors with very high energy resolution (approximately 30 meV) ARPES, which has allowed us to study the very low energy scale electronic structure. This low energy regime is by far the most interesting since the physics of the material (superconductivity, magnetism, etc.) is all directly a function of the near Fermi-surface electronic structure.

A. Fundamental studies of the superconducting energy gap

In these studies we have concentrated primarily on single crystals of $Bi_2Sr_2CaCu_2O_{8+d}$ (BSCCO) because these crystals cleave very nicely with little or no surface reconstruction and are very stable in the vacuum environment in which we must do our experiments. By taking photoemission spectra over a range of temperatures above and below the superconducting transition temperature (T_c), we are able to clearly observe the effects of the superconductivity on the electronic structure of BSCCO. As in conventional superconductors, we found that a gap (the superconducting energy gap) exists in the excitation spectrum of the high T_c superconductors. However, we also detected spectral weight changes that accompany this gap formation in BSCCO that are significantly different than those observed in conventional superconductors or predicted by any conventional theories of superconductivity [Dessau a]. In the conventional or BCS picture of superconductivity, only the very low energy electronic states (those within the gap value Δ of the Fermi energy) are affected by superconductivity. Our results show quite strong effects of the superconductivity at much higher energies (on the order of 4Δ, or approximately 90 meV). This is shown in Fig. 1, in which it is seen that along the G-M zone diagonal, there is a dip or depletion of spectral weight centered around -90 meV as the sample is cooled into the superconducting state. Along the G-X zone edge, the dip is very weak or not present. Another key result from this study is that the superconducting energy gap is larger along the zone diagonal than along the zone edge; that is, the superconducting energy gap is anisotropic in the $k_{x,y}$ plane of the Brillouin zone.

This work has garnered a great amount of interest from the theoretical physics community, with more than a few believing that the effects that we have discovered may be one of the keys to reaching an understanding of the microscopic mechanism of superconductivity in the high-temperature superconductors.

B. The metal/superconductor interface and the proximity effect.

, We have made use of the surface sensitivity of photoemission spectroscopy in our studies of the interfaces of the high temperature superconductors with other materials. These studies are technologically very important because any application of the high T_C superconductors will require that they come into contact with the outside world in some way. As will be shown, these studies have also uncovered some new and very exciting physics.

A major thrust of our program was a search for a proximity effect-induced superconducting energy gap in metallic overlayers on high temperature superconductors. The proximity effect is a phenomenon in which Cooper pairs (the superconducting charge carriers) will leak from a superconductor to a normal metal if the electrical contact between the two is good. The normal metal will then become a superconductor itself, that is, it will obtain the excitation spectrum of a superconductor. The study of the proximity effect between the high T_C superconductors and normal metals has both scientific and technological importance. For device work, a promising route is the construction of superconductor-normal-superconductor (SNS) proximity effect Josephson devices. From a scientific viewpoint, the coupling of the superconductivity to other materials will help us to better understand the nature and, hopefully, the origin of the superconductivity in the high T_C superconductors. The proximity effect has not, however, been unequivocally demonstrated to exist in metal/high T_C junctions. We have searched for the proximity effect in an original and unique way, which makes use of the surface sensitivity of photoemission spectroscopy [Dessau b]. We prepared junctions of the highest attainable quality in a number of different geometries, and then directly searched for the effects of a superconducting gap in the surface layer (gold) with high-resolution, temperature-dependent photoemission spectroscopy. We did not observe such an effect, but were able to place an upper limit of approximately 5 meV on its existence, which is well below the superconducting gap value of approximately 20-25 meV in the bulk high T_C superconductor.

A further set of experiments was carried out in which we slightly perturbed the outermost surface layer, allowing us to pin-point a number of effects as arising from either the surface layer, or from deeper in the bulk of the crystal [Wells]. We deposited sub-monolayer coverages of gold (~1/2 Å) on single crystals of BSCCO, at which point the gold overlayer was not yet metallic and so did not contribute to the states near the Fermi level. Along the Brillouin

33

zone edge (G-X), the gold caused no change to either the states at the Fermi level or to the gap structure in the superconducting state. However, along the Brillouin zone diagonal both of these features were destroyed when the gold overlayer was deposited. The first conclusion that can be drawn from this experiment is that for the undisturbed material, the superconducting gap exists undiminished up to the atomic surface. This is contrary to predictions made based on conventional 3-d Ginzburg-Landau theory and has strong implications for theories of the mechanism of high temperature superconductivity.

A second conclusion from this experiment has to do with the fact that the material consists of cation-oxygen layers. Only two types of layers in this material can possibly contribute electronic states at the Fermi level: the Bi-O layer and the Cu-O layer. Earlier experiments have shown that the Bi-O layer forms the cleavage plane while the first Cu-O layer is about 5Å deep [Lindberg]. Since it is highly unlikely that a Au surface layer could disrupt a deeper layer (Cu-O) without disrupting the surface layer (Bi-O), we can conclude that the region that was disrupted (the zone diagonal) had strong Bi-O character while the undisturbed region (the zone edge) is of almost purely Cu-O character. These are the atomic characters predicted by band theory.

We are continuing these ultra-high energy resolution ARPES experiments on the above materials systems, as well as branching out into the study of other systems, such as; the electron doped or n-type high T_c superconductors, buckyballs (C_{60}), and UHV in-situ grown and measured high T_c thin films.

References

[Dessau a] D.S. Dessau, B.O. Wells, Z.-X. Shen, W.E. Spicer, R.S. List, A.J. Arko, D.B. Mitzi, and A. Kapitulnik, *Physical Review Letters* **66**, 2160 (1991).

[Dessau b] D.S. Dessau, B.O. Wells, Z.-X. Shen, W.E. Spicer, R.S. List, A.J. Arko, C.B. Eom, D.B. Mitzi, A. Kapitulnik and T.H. Geballe, *Applied Physics Letters* **58**, 1332 (1991).

[Lindberg] P.A.P. Lindberg, Z.-X. Shen, B.O. Wells, D.S. Dessau, D.R. Mitzi, I. Lindau, W.E. Spicer and A. Kapitulnik, *Phys. Rev. B* **39**, 2890 (1989).

[Wells]   B.O. Wells, Z.-X. Shen, D.S. Dessau, W.E. Spicer, C. G. Olson, D.B. Mitzi, A. Kapitulnik, R.S. List and A.J. Arko, *Phys. Rev. Lett.* **65**, 3056 (1990).

JSEP Supported Publications

1. D. S. Dessau, B. O. Wells, Z.-X. Shen, W. E. Spicer, R. S. List, A. J. Arko, D. B. Mitzi, and A. Kapitulnik, "Anomalous spectral weight transfer at the superconducting transition of $Bi_2Sr_2CaCu_2O_{8+d}$" *Physical Review Letters* **66**, 2160 (1991).
2. D. S. Dessau, B. O. Wells, Z.-X. Shen, W. E. Spicer, R. S. List, A. J. Arko, C. B. Eom, D. B. Mitzi, A. Kapitulnik and T. H. Geballe, "Search for a proximity effect induced gap in gold/high-Tc junctions" *Applied Physics Letters* **58**, 1332 (1991).
3. Z.-X. Shen, D. S. Dessau, B. O. Wells, C. G. Olson, D. B. Mitzi, L. Lombardo, R. S. List and A. J. Arko, "Evidence of chemical-potential shift with hole doping in $Bi_2Sr_2CaCu_2O_{8+d}$" *Physical Review* B **44**, 12098 (1991).
4. D. S. Dessau, Z.-X. Shen, B. O. Wells, W. E. Spicer, A. J. Arko, "Spectral weight changes at the superconducting transition of $Bi_2Sr_2CaCu_2O_{8+d}$" Proceedings of the Workshop on Fermiology of High-Tc Superconductors, *Journal of Physics and Chemistry of Solids* **52**, No. 11/12, 1401 (1991).
5. Z.-X. Shen, D.S. Dessau and B.O. Wells, "Electronic structure and the superconducting gap of $Bi_2Sr_2CaCu_2O_{8+d}$" Proceedings of the University of Miami workshop on "Electronic Structure and Mechanisms for High-Tc Superconductivity" (1991).
6. D. S. Dessau, Z.-X. Shen, B. O. Wells, W. E. Spicer, R. S. List, A. J. Arko, C. G. Olson, D. B. Mitzi, C. B. Eom, A. Kapitulnik and T. H. Geballe, "The gold/high-temperature superconductor interface; metallicity of the near surface region and a search for the proximity effect" *Journal of Vacuum Science Technology* **A9**, 383 (1991)
7. B. O. Wells, Z.-X. Shen, D. S. Dessau, W. E. Spicer, Reply to 'Anomalous enhancement of $Bi_2Sr_2CaCu_2O_8$ Fermi-level states near the O 2s threshold' *Phys. Rev.* B **44**, 882 (1991).
8. Z.-X. Shen, R. S. List, D. S. Dessau, B. O. Wells, A. J. Arko, R. Bartlett, O. Jepsen, F. Parmigiani, I. Lindau, and W. E. Spicer, "Electronic structure of NiO --- correlation and band effects" *Phys. Rev.* B **44**, 3604 (1991).

9. Z.-X. Shen, R. S. List, D. S. Dessau, A. J. Arko, R. Bartlett, O. Jepsen, B. O. Wells and F. Parmigiani, "Angle-resolved photoemission of NiO (001)" *Solid State Communication* **79**, 623 (1991).

10. B. O. Wells, Z.-X. Shen, D. S. Dessau, W. E. Spicer, C. G. Olson, D. B. Mitzi, A. Kapitulnik, R. S. List and A. J. Arko, "Test of the Fermi surface in $Bi_2Sr_2CaCu_2O_8$ -- Metallicity of the Bi-O layer" *Phys. Rev. Lett.* **65**, 3056 (1990).

11. D. S. Dessau, R. S. List, Z. X. Shen, B. O. Wells, W. E. Spicer, R. S. List , A. J. Arko, R. J. Bartlett, Z. Fisk, S-W. Cheong, D. B. Mitzi, A. Kapitulnik and J. E. Schirber, "Electronic structure of the gold/$Bi_2Sr_2CaCu_2O_8$ and gold/$EuBa_2Cu_3O_{7-d}$ interfaces as studied by photoemission spectroscopy" *Applied Physics Letters* **57**, 307 (1990)

12. Z.-X. Shen, R. S. List, D. S. Dessau, F. Parmigiani, A. J. Arko, R. Bartlett, B. O. Wells, I. Lindau, and W. E. Spicer, "Resonance photoemission from CuO and Cu2O single crystals" *Phys. Rev.* B **42**, 8081 (1990).

13. Z.-X. Shen, J. W. Allen, P. A. P. Lindberg, J. S. Kang, W. Ellis, D. S. Dessau, B. O. Wells, A. Borg, I. Lindau, and W. E. Spicer, "Photoemission study of CoO" *Phys. Rev.* B **42**, 1817 (1990).

14. P. A. P. Lindberg, B. O. Wells, Z.-X. Shen, D. S. Dessau, D. B. Mitzi, I. Lindau, W. E. Spicer and A. Kapitulnik, "Interaction of overlayers of Al and Rb with single crystalline surfaces of $Bi_2Sr_2CaCu_2O_8$" *J. of Appl. Phys.* **67**, 2667 (1990).

15. Z.-X. Shen, P. A. P. Lindberg, B. O. Wells, D. S. Dessau, A. Borg, I. Lindau, W. E. Spicer, W. P. Ellis, G. H. Kwei, K. C. Ott, J.-S. Kang and J. W. Allen, "Electronic structure of monoclinic $BaBiO_3$" *AVS/AIP conference proceedings* no. 200, **30** (1990).

16. B. O. Wells, P. A. P. Lindberg, Z.-X. Shen, D. S. Dessau, W. E. Spicer, I. Lindau, D. B. Mitzi and A. Kapitulnik,"Nature of the valence band states in $Bi_2(Ca,Sr, La)_3Cu_2O_8$." *AIP/AVS Conference Proceedings* no.200, **36** (1990).

17. O. Gunnarsson, J. W. Allen, O. Jepsen, T. Fujiwara, O. K. Anderson, C. G. Olsen, M. B. Maple, J.-S. Kang, L. Z. Liu, J.-H. Park, R. O. Anderson, W. P. Ellis. R. Liu, J. T. Market, Y. Dalichaouch, Z.-X. Shen, P. A. P. Lindberg, B. O. Wells, D. S. Dessau, A. Borg, I. Lindau, and W. E. Spicer, "Resonance photoemission for Nd_2CuO_4" *Phys. Rev.* **B41**, 4811 (1990).

Unit: 5

TITLE : Semiconductor Laser Structures
for Optical Interconnects

PRINCIPAL INVESTIGATOR : S. S. Wong

GRADUATE STUDENT : S. Biellak

Scientific Objectives :
The objective of this work is to study the interplay between geometry and physics in a semiconductor laser in order to achieve high efficiency optical emitter structures for applications in optical data communication.

Summary of Research :
As integrated systems become increasingly complex, a limiting factor in the performance is the rate at which data can be communicated within the systems. Optical links can potentially deliver the high data rate required. The goal of this project is to design, fabricate and characterize broad-area, high power, diffraction-limited semiconductor laser diodes by making use of an unstable-resonator-based optical design. Diode lasers with this level of output power and this quality of optical beam output are needed for applications to high-capacity optical data links in computers and communications, free space optical communications, high-speed read-write optical disks, and diode-pumping of solid-state lasers and wavelength doublers for video and projection display systems.

Typically, the width of a laser diode is constrained to be on the order of tens of microns if it is to operate with a single lateral mode (i.e., diffraction limited). Beyond this width, internal self-focusing, or "filamentation", effects force multimode operation. Consequently, single stripe AlGaAs/GaAs lasers are only capable of roughly one hundred milliwatts continuous-wave (CW) output power in a diffraction-limited spot. Commercially available multi-watt CW diode laser systems suffer from poor, multi-mode output beam quality due to the large numbers of individual lasing elements employed.

37

The concept of "unstable optical resonator" was developed a number of years ago and is now widely used in conventional laser devices of many different types. By using divergent rather than convergent mirrors in the laser cavity, optical rays bouncing back and forth in such a cavity will exhibit an exponential divergence outward from the axis. Rather than executing stable periodic oscillations back and forth about the cavity axis on repeated bounces as in a stable periodic focusing system, rays in an unstable cavity diverge exponentially outward from the cavity axis on repeated bounces. Despite this geometrically unstable behavior for optical rays, when the diffractive properties of optical waves are taken into account, an unstable cavity of this sort still exhibits a perfectly stable set of resonant modes or lateral cavity modes which can be used as the basis of laser oscillation. These modes extend in general across the full width of the unstable cavity, and tend to have somewhat larger output coupling than the modes in high-reflectivity stable laser resonators. There is also typically a sizable discrimination between the losses of the lowest-order and the next higher-order lateral modes in an unstable resonator. As a result of this, a laser device using an unstable resonator tends to oscillate only in the single lowest-order lateral mode of the cavity. This in turn leads to a single-mode or diffraction-limited output beam. Unstable cavity modes are thus very well suited to extracting high power from the full width of a wide-stripe diode laser. Our work investigates the means of fabricating unstable optical cavities in AlGaAs/GaAs single-quantum-well heterostructures. The simplest method is to etch a curved, concave mirror facet at one end of the optical cavity, effecting beam divergence upon internal reflection of the circulating field. Previous groups' efforts to do this via wet chemical or focused ion beam etching have produced unreliable facets and mediocre beam quality [Salzman] [Tilton]. We have developed a reactive-ion-etch process using $SiCl_4$ that consistently achieves curved mirrors facets with a surface roughness less than 1/5 of the wavelength (in the semiconductor) over a wide range of mirror magnifications. A scanning electron micrograph of an etched facet is shown in Fig. 1a. We are now refining our processes to allow for the fabrication and testing of these laser diodes with a wide variety of cavity configurations (i.e., length, width, and radius of curvature). A photograph of several of the etched unstable resonator cavities is shown in Fig. 1b.

(a)   (b)
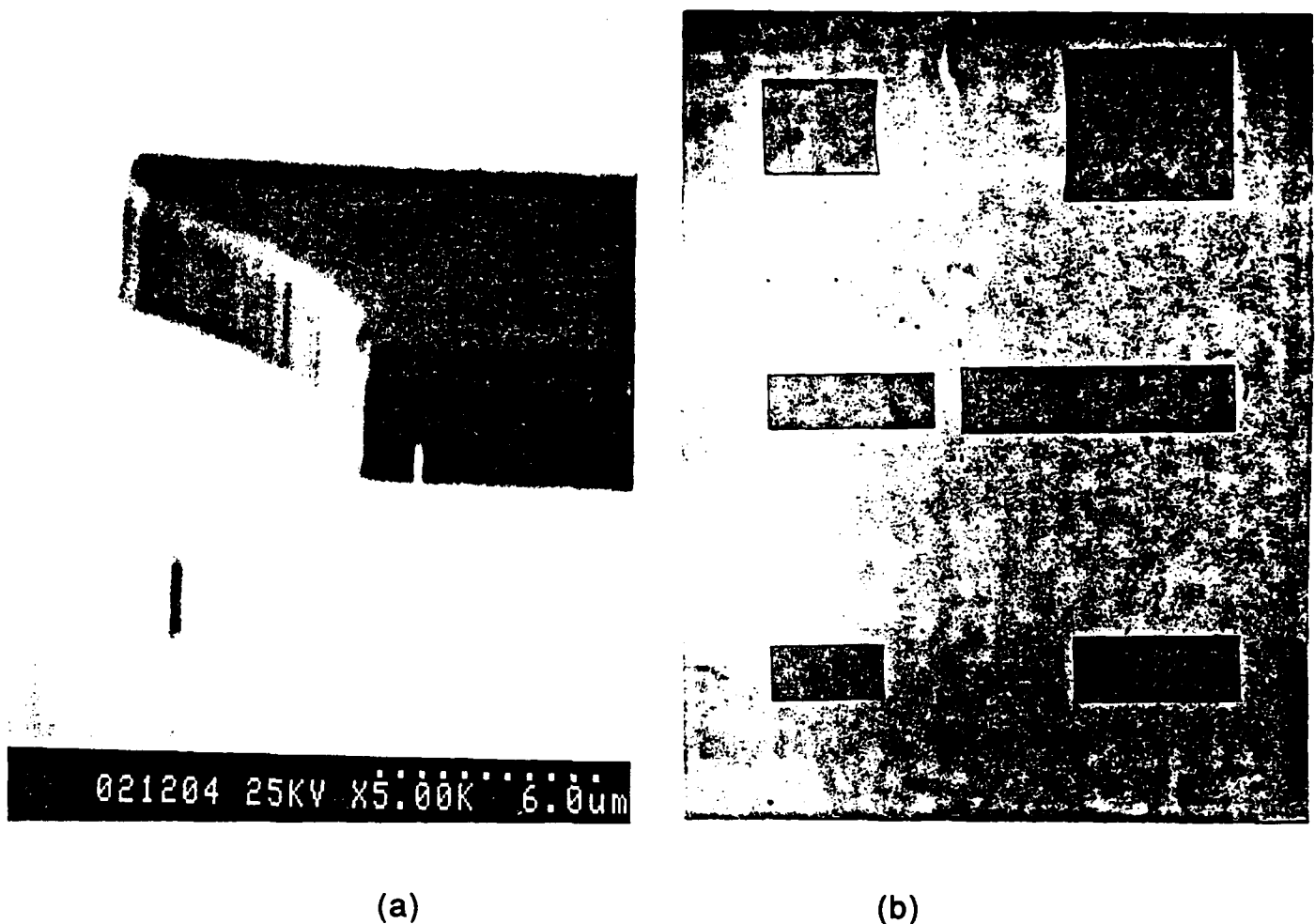
Figure 1 (a) SEM photograph of a dry-etched curved facet, and (b) photograph of various cavities; the radius of curvature is 800 μm for the devices on the left and 200 μm for those on the right.

Future Directions

We expect to complete the fabrication of the first set of laser diodes with curved mirror facets in the near future. The performance behavior of these devices will be measured and the quality of the facets will be determined. Concurrently, we have been performing a series of numerical simulations examining an alternative, cleaved facets unstable-resonator configuration, where the cavity field divergence is achieved using a quasi-Fresnel lens formed by etching periodic steps into the diode cladding layer. We expect to begin fabricating such "unstable lens-guide" devices in the near future. Finally, we hope to monolithically integrate one of these laser diodes with a collimating mirror and grated amplifier section to produce multi-watt CW surface-emitting output.

References

[Salzman] J. Salzman, et al., *Appl. Phys. Lett.*, **46**, p. 218, 1985.

[Tilton] M. L. Tilton, et al., *IEEE J. Quantum Electron.*, **27**, p. 2098, 1991.

UNIT: 6

TITLE: Quantum Computing

PRINCIPAL INVESTIGATOR: J. D. Plummer

GRADUATE STUDENT: B. Biegel

Scientific   Objectives:

The objectives of this research are three-fold. First, we hope to contribute to the understanding of ways in which quantum effects can be used to facilitate the operation of electronic devices which are potentially much more efficient and highly integrable than conventional electronic devices.   Second, and central to this research, we will develop a numerical simulation tool for the simulation of these quantum electronic devices.   Finally, we will test this simulator's predictions by fabricating and testing our own SiGe-based quantum devices and by comparing its predictions with published measurements of the operation of III-V quantum devices.

1. Summary of Research:

The basic direction of knowledge flow in this research project is from the more general to the more specific: from theory to simulation to experiment.   Our theoretical investigations of quantum computing serve to guide our simulation efforts, which in turn guide our experimental efforts. For this reason, we will present the results of our work in this order. Because the central objective of this work is the development of an accurate quantum device simulator, more space will be given in this report to progress made towards this objective.

1.1 Theory:

Quantum computing (QC) is the concept of producing useful computing (digital or analog signal processing) using integrated electronic devices whose operation is based fundamentally on quantum effects.   At present, much of the theory of how we might produce useful computing from quantum effect devices remains unknown, or at least uncertain.   For this reason, during the past year, we have continued our effort to synthesize and extend the theory of quantum computing (see previous annual report and [Randall], [Bate

86], and [Bate 87] for more information). This work provides both motivation and direction for our simulation and experimental work.

Recall that the basic motivation for quantum computing is that it will be the only option for continued device scaling beyond the "quantum barrier" faced by conventional electronics [Bate 88]. Defining motivation is the first, albeit the most superficial, part of quantum computing theory. However, it compels us to build a deeper understanding. Our efforts in this regard focused first on *how* to efficiently develop such a revolutionary new technology. The realization of quantum computing's great potential will not be easy: in addition to the great technological challenge of simply *building* quantum scale integrated circuits (QSICs), we are conceptually challenged to use completely different phenomena to control the operation of electronic devices at unprecedented integration densities. Quantum computing may only become a reality with our ability to look at computing in a new way: with our present knowledge, but without our usual preconceptions. In particular, in our development of QC, we must:

- use ULSI as both a starting point (as a knowledge and technology base) and a yardstick (since it embodies our current state-of-the-art in computing),
- use the "optical analogy" (the similarity between optical and quantum electronic systems) as another excellent source of knowledge, concepts, and insight for use in the development of QC, especially with the advances in all-optical computing [Bell],
- challenge ourselves to take advantage of the wave nature of carriers where it serves us best (especially at the device level) and the particle nature where it is best,
- use limitations of these technologies to suggest solutions for QC, as was done in the initial proposal of QC, and
- re-evaluate each conventional rule of computing systems to determine whether or not it is fundamental and applicable to QC.

Our theoretical investigation is based on this philosophy. The optical analogy [Gaylord] to quantum computing systems was further explored during this period, to aid in the development of QC science. By virtue of the fact that a QC system is a wave system, some basic wave effects will automatically occur, including wave-like propagation and interference. To produce other basic wave phenomena, including reflection, refraction, and diffraction,

physical structures which cause reflection and refraction are needed. Both of these effects can be produced by changes between adjacent materials either in the energy band minimum (band offsets, or heterojunctions), or the effective mass. It happens that band offsets are easier to produce and more effective in creating desired wave effects than are effective mass changes. From this fact arises the concept of band gap engineering [Capasso 87], which is used to design quantum, optoelectronic, and other heterojunction devices. Thus, band offsets are usually used to design the function of quantum devices, but the effective mass determines their relative size: devices scale as $m^{(-1/2)}$. Thus, structures in SiGe quantum devices will be 2-3 times smaller than in III-V quantum devices. These and other conclusions concerning the optical analogy are included in an extensive review of the theory of quantum computing that is under development (for inclusion in a thesis and possible publication).

1.2 Simulation:

Our theoretical investigation of quantum computing addresses the more general issues, such as what characteristics a quantum computing system will have, and what characteristics quantum computing itself will have. On the other hand, our efforts to develop a quantum simulation capability answers the need to examine in more detail particular quantum phenomena and devices, to determine accurately how quantum systems will operate and thus whether and how we can extract useful computing from them. To confirm the predictions and test the accuracy of our simulation tool, we will fabricate simple quantum effect devices in the SiGe material system as discussed in Section 1.3. We will also compare its predictions for GaAs-based structures with experimental data from other researchers.

Our main quantum simulation tool is based on solving the quantum transport equation for the Wigner function (WF) [Frensley] [Jensen 90]. This approach is the quantum mechanical analog of solving the Boltzmann transport equation, from which are derived all mathematical models used to simulate conventional (drift-diffusion) electronic systems. The resonant tunneling diode (RTD) [Capasso 86] serves as the prototype quantum device to be modeled in our simulations (and fabricated experimentally; see Section 1.3) for reasons discussed in our previous annual report. All particular results given in this report are for p-type SiGe RTDs, since SiGe is our chosen heterojunction material system. At the conclusion of the

previous reporting period, we had written and performed initial testing on both a steady-state and a transient WF simulation program. Both gave reasonable predictions, based on knowledge and experimental results in the literature, of the operation of both SiGe-based and GaAs-based RTDs. For details, see the previous annual report.

During the past year, several improvements and optimizations have been made to our WF quantum device simulator. Despite its inherent advantages [Jensen 91], the WF simulation method is relatively expensive in terms of time and memory requirements compared to other possible quantum simulation methods, so that optimizing its execution is a very important concern. The most significant changes made and their results are summarized in Table 1 for a typical simulation. The first change was a result of our decision to convert from the Fortran to the C program language. Among other advantages, C allows easier communication with the kinds of tools one needs to create a widely useful simulation tool: the operating system, shell, X windows and graphical interfaces, and other C programs on which UNIX platforms are based. Further, C compilers are generally better at optimizing programs than Fortran compilers. Thus, we realized a 20% decrease in typical simulation times through the conversion to C. During this process, we also discovered a method we call windowed Gaussian elimination (WGE), which allowed us to decrease the memory requirement of the simulator by almost 50% (or alternately, increase the number of grid nodes). This speeded up execution by an additional 15%. Finally, by moving to a more capable workstation, simulation time was decreased to about 1/3 of the original.

Table 1: WF Simulator Time and Memory Improvements

Workstation	Language	WGE	Time (s) / %	Memory (MB)
DEC 3100	Fortran	no	194.2 / 100.0	16.31
DEC 3100	C	no	155.1 / 79.9	16.44
DEC 3100	C	yes	125.2 / 64.5	8.86
DEC 5000	C	yes	67.0 / 34.5	8.86

In several respects, our WF simulation program was inadequate or deficient. Its shortcomings include its relatively large time and memory requirements, and its current lack of numerical robustness (although we are continually working to improve this), resulting in unphysical results for some structures, and our lack of certainty in

its predictions. Because of these concerns, we added a second tool, based on the transmission matrix (TM) method [Tsu], to our quantum device simulation suite. The TM method is inherently less calculation intensive, faster, and more robust. However, it requires a much greater abstraction of the carriers in the system [Jensen 91], with the result that certain important effects, such as scattering and dynamic boundary conditions, can not be meaningfully implemented in this approach. The current implementation of the WF approach does not include such effects yet (see Section 2), so the TM method can serve to check the WF simulation results. Even when these effects are implemented in the WF simulator, relatively quick TM simulations can serve to narrow the range of relatively expensive WF simulations required. Finally, comparison of TM and WF simulations will indicate how we can improve our (main) WF simulation tool, and will hopefully give us confirmation of, and confidence in, our WF simulations.

Figure 1 shows the simulated I-V characteristic, from both WF and TM simulations, for the same SiGe RTD structure. We can see both similarities and differences in the two results. Both simulation methods predict currents of the same order of magnitude, indicate a relatively strong resonance condition at an applied bias of about 0.15 V, an anti-resonance at roughly 0.2 V, and show evidence of a weak resonance near 0.75 V applied bias. However, the details of the two predictions are quite different. The TM method can give unphysically sharp negative differential resistance (NDR) transitions and prolonged anti-resonant regions. This is due to the TM method's treatment of carriers at a given energy as perfectly in phase, but completely independent of carriers at other energies. The WF method does not abstract phase information in this way, the result of which is a less dramatic, and supposedly more accurate, I-V prediction. Thus, our two quantum device simulation approaches show both the similarities and the differences that we expect. Again, we should note the primary role of the WF simulator and the secondary role of the TM simulator.

The chart contains the following annotations:

T-Matrix
Simulation

Wigner Function
Simulation

T-Matrix Simulation:
Peak current: 2.19×10^5 A/cm^2
Peak-to-valley ratio: 24.75
Peak voltage: 0.149 V
Valley voltage: 0.198 V

Wigner Function Simulation:
Peak current: 1.32×10^5 A/cm^2
Peak-to-valley ratio: 1.16
Peak voltage: 0.146 V
Valley voltage: 0.194 V

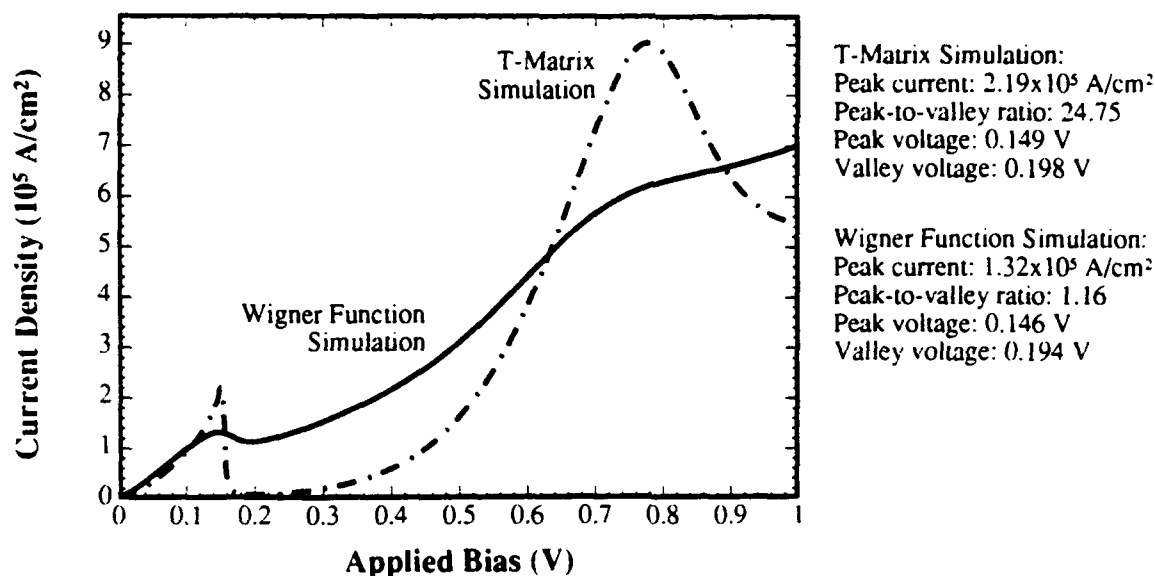Axis labels: Current Density (10^5 A/cm^2) (vertical), Applied Bias (V) (horizontal)

Figure 1: Simulated SiGe RTD I-V Characteristic

Of course, the code of both the WF and TM simulators is under constant revision to add capabilities, versatility, and robustness. The main enhancements completed during the current reporting period were:

- Merging the common parts of the two simulation methods, such as input functions, output and plotting routines, and band offset and energy band profile calculations. By merging code, future enhancements (such as self-consistency) will be immediately available to both simulators.
- A new method for calculating band offsets which allows arbitrary band offset models for any material or material system to be easily added. For example, the SiGe system includes lattice constant (i.e., stress) dependence, while the (Al,Ga,In)As system does not.
- A new method for estimating the self-consistent energy band profile, which should prove to be a reasonable initial guess when full self-consistency is implemented.
- Many output enhancements, particularly in terms of output plots and plotting capabilities, allowing more complete and precise analysis of the simulation results.

A final reason for our interest in the TM simulation approach is its possible use to extend the WF approach to more dimensions. 2-D and 3-D simulation capability will be needed with quantum devices just as they are being pursued for conventional devices. At

a minimum, using the TM approach alone, it should be feasible to simulate 2-D and perhaps 3-D problems.

1.3 Experiment:

Our final objective in this project is to experimentally investigate particular quantum-scale structures using the SiGe material system. Perhaps the most important reason for our choice of the RTD as a simulation device is that it is possible to fabricate and test these devices with existing technology. Thus, we can compare simulated and measured data for the same structure, and meaningfully assess the accuracy of our simulation tool. Our experimental work began during the current reporting period in collaboration with K. L. Wang's group at UCLA [Chang]. The specified SiGe material layers will be grown by MBE at UCLA, while the actual device fabrication will be done at Stanford in the Center for Integrated Systems' (CIS) Integrated Circuits Lab (ICL).

During the past year, we made our first experimental measurements on SiGe RTDs supplied by UCLA. Although in the future, we plan to complete the device processing in CISs ICL, these pre-packaged devices were provided by UCLA to allow us to prepare and test a measurement station for room temperature and low temperature (77 K and 4 K) measurements. We were able to use an HP 4145 test station set up by J. S. Harris' group at Stanford [Chou] for this purpose. Figure 2 shows a measured I-V curve at 77 K for one such SiGe RTD. It also shows the TM simulated I-V curves for the device (WF simulation not shown for reasons discussed below). Many of the noteworthy features of this plot are a result of the measured RTD having a relatively thick tunnel barrier (5.5 nm, versus 1.5 nm in the samples simulated in Figure 1). This results in an impractically low current density and very sharp resonant states. Because the TM method is relatively fast and uses little computer memory, it is feasible to use a very fine energy resolution, so that we have a chance of finding such sharp resonances (dotted curve would result with infinite energy resolution). With the WF method, such fine resolution is not feasible, with the result that these very sharp resonances were missed entirely. In this extreme of thick barriers, performance has actually been degraded. Practically, there is no benefit for making the resonant width much narrower than the energy spread of the incident carriers. (about 40 meV in this case).
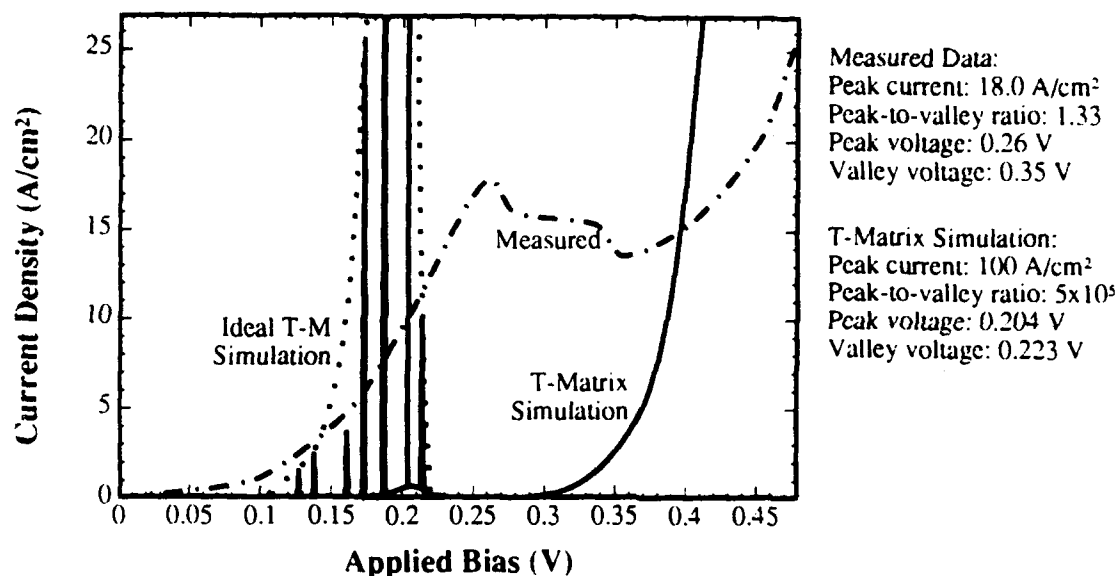
Figure 2: Measured SiGe RTD I-V Characteristic

Three features of the measured I-V curve in Fig. 2 are a result of external effects: the higher than predicted voltage at resonance, the plateau in the NDR region, and the leakage current. The measured resonance voltage will typically be higher than the simulated value due to parasitic series resistances, such as contact resistance. The plateau in the NDR region of the measured I-V curve is the result of external measurement circuit oscillations [Young]. Finally, the leakage current (total current minus ideal resonance current) is a result of perimeter currents and inelastic tunneling currents, which are non-negligible at these low current levels. In general, we must contend with non-zero parasitics and non-ideal external circuit elements in the measurement set-up, with the result that the ideal (internal) simulation of quantum devices will never exactly match the measured results. However, by incorporating the WF or TM simulation results in some way (either I-V characteristics directly or lumped-parameter "equivalent circuits") into a circuit simulator, we can achieve a meaningful comparison. Merging quantum device and circuit simulators is thus one of our long-term goals (see Section 2). Incorporating quantum device simulation results into circuit simulators would also allow us determine how independent quantum devices would function and interact in a circuit.

After completing our initial RTD measurements, we needed to decide on particular structures to have grown at UCLA. We realized that there are two important constraints on the layer thicknesses and germanium content of the RTD layers. The first is a result of

the fact that SiGe is a strained material system, and therefore that layer widths must be constrained in order to maintain defect free (i.e., unrelaxed) interfaces and layers. The second constraint is that we require a minimum band offset of at least 0.2 eV to achieve useful RTD performance. These constraints were combined, using the [Van de Walle] model of SiGe band offsets and the [Bean] data for SiGe layer critical thickness, to derive SiGe RTD design space. Using this tool and other considerations, we proposed a matrix of RTD layer structures for MBE growth at UCLA. This work is currently underway, upon completion of which we will conduct the device fabrication and subsequent measurement here at Stanford.

2.   Goals:

In the future, this work will continue to evolve from theory to simulation to experiment. We will continue our theoretical investigation of quantum computing, since insight and analysis in this area will indicate new directions to pursue in simulation and experiment.

Work on our quantum device simulation programs will continue as well. Through the simulator development done during this period, we have made progress toward the realization of most of our long-term goals (see the previous annual report). Briefly, these include:

- Full self-consistency. (WF and TM)
- Dissipative effects. (WF)
- Numerical robustness. (WF and TM)
- More physical boundary conditions. (WF and TM)
- Equivalent circuit generation and circuit simulation. (WF and TM)
- 2-D and 3-D capability. (WF and TM)

Our experimental work to fabricate and test SiGe RTDs and related quantum-scale structures will continue in parallel with our simulation effort, in order to test the simulator's predictions, and to investigate the SiGe material system as a heterojunction material alternative. Of course, the ultimate goal of creating a simulator is to guide (much more expensive) experiment. Of the above planned improvements to our quantum device simulator, we have found that at a minimum, self-consistency and circuit simulation must be implemented before simulation results can be accurately compared with experimental measurements. This finding confirms our belief that, during simulator development, experimental data serves as an essential monitor on the accuracy of

the simulation tool. Thus, we will continue to actively pursue the experimental phase of our work.

JSEP Supported Publications:

References:

[Bate 86]   R. T. Bate, *Superlattices and Microstructures*, **2**, 9, (1986).

[Bate 87]   R. T. Bate, G. A. Frazier, W. R. Frensley, J. W. Lee, and M. A. Reed, *Proceedings of the SPIE, Vol. 792: Qauntum Well and Superlattice Physics*, p. 26, 1987.

[Bate 88]   R. T. Bate, *Scientific American*, **258**, 96, (1988).

[Bean]   J. C. Bean, L. C. Feldman, A. T. Fiory, and I. K. Robinson, *J. of Vac. Sci. and Tech.*, A2, 436, (1984).

[Bell]   T. E. Bell, *IEEE Spectrum*, **23**, 34, (1986).

[Capasso 86]   F. Capasso, K. Mohammed, and A. Y. Cho, *IEEE Journal of Quantum Electronics*, **22**, 1853, (1986).

[Capasso 87]   F. Capasso., *Science*, **235**, 172, (1987).

[Chang]   S. J. Chang, C. F. Huang, M. A. Kallel, K. L. Wang, Jr. R. C. Bowman, and P. M. Adams, *Appl. Phys. Lett.*, **53**, 1835, (1988).

[Chou]   S. Y. Chou and Jr. J. S. Harris, *Applied Physics Letters*, **52**, 1422, (1988).

[Frensley]   W. R. Frensley, *Physical Review*, **B36**, 1570, (1987).

[Gaylord]   T. K. Gaylord and K. F. Brennan, *Journal of Applied Physics*, **65**, 814, (1989).

[Jensen 90]   K. L. Jensen and F. A. Buot. *Journal of Applied Physics*, **67**, 7602, (1990).

[Jensen 91]   K. L. Jensen and A. K. Ganguly, *Technical Digest of the IVMC 91*, p. 18, 1991.

[Randall]   J. N. Randall, M. A. Reed, and G. A. Frazier, *Journal of Vacuum Science and Technology* **B7**,1398, (1989).

[Tsu]   R. Tsu and L. Esaki, *Appl. Phys. Lett.*, **22**, 562, (1973).

[Van de Walle] C. G. Van de Walle and R. M. Martin, *Physical Review*, **B34**, 5621, (1986).

[Young]   J. F. Young, B. M. Wood, H. C. Liu, M. Buchanan, D. Landheer, A. J. SpringThorpe, and P. Mandeville, *Applied Physics Letters*, **52**, 1398, (1988).

Unit: 7

TITLE: Applications of Polycrystalline Silicon-Germanium Films in MOS Technologies

PRINCIPAL INVESTIGATOR: K. C. Saraswat

GRADUATE STUDENT: T.-J. King

Scientific Objective

The objective of this work is to characterize the physical and electrical properties of polycrystalline silicon-germanium (poly-SiGe) films and to assess their advantages over poly-Si for various applications in MOS technologies. In the current reporting period, we have developed a new technology to fabricate MOS thin-film transistors (TFTs) in poly-SiGe films. Compared to a poly-Si TFT technology, an optimized poly-SiGe TFT technology should be able to provide a lower-temperature, shorter-time processing capability at no expense to performance. Such a technology would be useful for a variety of applications, such as; large-area display driver and 3-D integrated circuits, and high-density SRAMs (as PMOS load transistors) because of its low-temperature processing advantage over a poly-Si TFT technology.

Summary of Research

Thin-film transistors (TFTs) can be used to integrate peripheral drive circuits with active matrix liquid-crystal displays (LCDs) on large-area substrates [Ohwada]. This integration greatly reduces the number of external connections and the number of external driver chips, thereby lowering cost and improving reliability. Low-temperature, short-time processing is essential in order to fabricate low-cost integrated circuits for large-area electronics. Hydrogenated amorphous silicon (α-Si:H) TFTs can be fabricated on glass substrates, but their low effective mobilities make them unsuitable for use in driver circuits for large-area displays. Polycrystalline silicon (poly-Si) TFTs have much higher effective mobilities (>10 $cm^2/V.s$); however, using conventional processing equipment, their fabrication requires higher temperatures (>600°C) and long-time processing, making them marginally compatible with large-area glass substrates.

We had previously demonstrated that p+ polycrystalline silicon-germanium (poly-SiGe) can be used as a variable-work-function CMOS gate electrode material [King1] and that PMOS TFTs with good device characteristics can be fabricated in poly-SiGe films [King2]. We have recently developed a new low-temperature poly-SiGe CMOS TFT technology for large-area electronics applications. Because the process temperatures never exceed 550°C, this technology would allow the fabrication of high-performance polycrystalline TFTs on low-cost glass substrates, without introducing any additional fabrication process complexity in comparison to α-Si TFT technology.

Properties of Poly-SiGe Films

Since the melting point of SiGe is lower than that of Si, physical phenomena controlling fabrication processes, such as; deposition, crystallization, grain growth, dopant activation, etc. occur at lower temperatures for SiGe than for Si. Poly-SiGe films with Ge mole fractions up to 0.6 are completely compatible with standard VLSI fabrication processes, and can be routinely deposited in commercially available LPCVD equipment, using silane (SiH_4) and germane (GeH_4) as the gaseous deposition sources [King1]. The Ge mole fraction in the deposited films varies with the percentage of GeH_4 in the gaseous deposition source and also with the deposition temperature. The percentage of GeH4 in the gaseous source can be varied linearly with temperature to obtain films of a specific Ge mole fraction. Depending upon their Ge mole fraction, the deposited films can be polycrystalline for deposition temperatures as low as 400°C (Fig. 1). By low-temperature (550°C) annealing, amorphous SiGe films can be crystallized to obtain large (~0.25 μm) grains. Poly-SiGe films are easily patterned using a conventional Si plasma etch process, and they can be doped by ion implantation. Anneal temperatures as low as 500°C can be used to achieve a high degree of dopant activation and thereby achieve low resistivity [King1]. The grain boundaries in poly-SiGe films can be passivated by hydrogen implantation followed by low-temperature (275°C) annealing [King2]. All of the aforementioned properties are conducive to a low-thermal-budget polycrystalline TFT technology. Along with the potential enhancement of carrier mobilities in SiGe [Pearsall], this provided the motivation for the investigation of the characteristics of low-temperature-processed poly-SiGe TFTs.
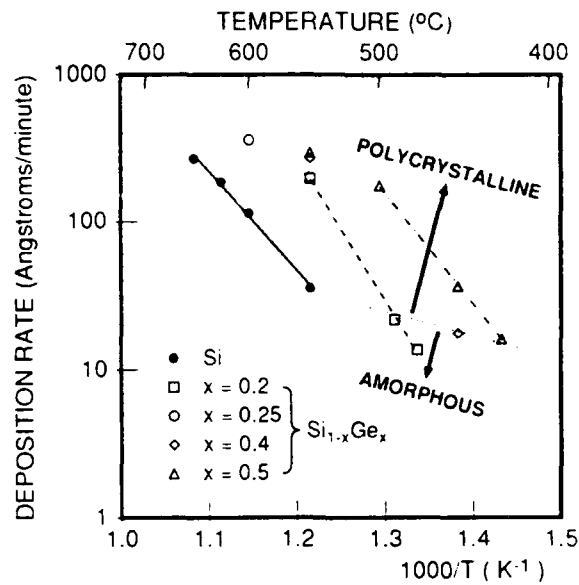
Figure 1   Deposition rate of $Si_{1-x}Ge_x$ films as a function of inverse temperature, for films of various Ge mole fractions. The dotted line delineates transition temperatures at which the deposited films begin to deposit in polycrystalline form.
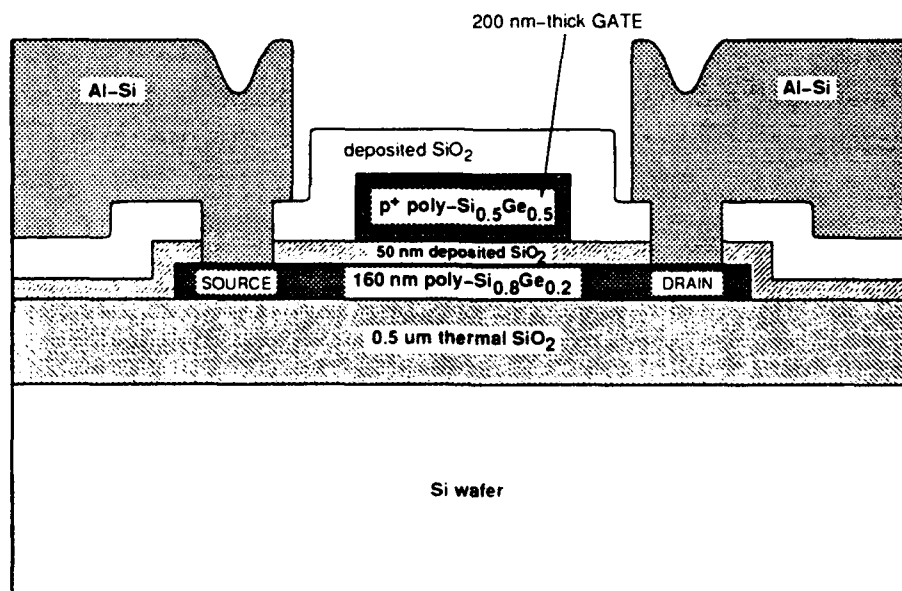

Figure 2   Schematic cross-sectional view of the thin-film transistors fabricated in this work.

Transistor Fabrication

Figure 2 shows a schematic cross-sectional view of the TFTs which were fabricated in this work. Standard self-aligned gate n- and p-MOSFET structures were fabricated using conventional Si processing techniques. Silicon wafers with 0.5 μm of thermally grown SiO_2 were used for the starting substrates. The 160 nm-thick active layer of $Si_{0.8}Ge_{0.2}$ was deposited (in amorphous form) at 490°C and then crystallized during a 24-hour anneal at 550°C. After the active regions were defined by plasma etching, a 50 nm-thick gate oxide was deposited by LPCVD at 400°C and was subsequently densified and annealed for 12 hours at 550°C. Poly-$Si_{0.5}Ge_{0.5}$ deposited at 500°C was used for the gate electrode material. This gate material was 200 nm thick, and was doped by boron implantation (4 x 10^{15} cm^{-2} @ 40 keV) followed by annealing for 1 hour at 550°C. After the patterning of the gate material, the source and drain regions were formed by ion implantation of phosphorus (4 x 10^{15} cm^{-2} @ 120 keV) and boron (4 x 10^{15} cm^{-2} @ 46 keV) for the n- and p-channel TFTs, respectively. A 1-hour anneal at 550°C was sufficient to achieve a sheet resistivity of 900 Ω/square in the PMOS source and drain regions; however, because of the high-energy and high-dose phosphorus implant conditions, the NMOS devices required ~60 hours @ 550°C in order to recrystallize the heavily implanted regions. This problem could have been easily avoided by using a lower implant dose and a lower implant energy for the n-channel devices. The final sheet resistivity of the p+ poly-$Si_{0.5}Ge_{0.5}$ gate material was ~200 Ω/square. Following the low-temperature SiO_2 (LTO) passivation-layer deposition, contact definition and Al-Si metallization steps, the devices were given a 30-minute forming-gas anneal at 300°C. Finally, the channel poly-$Si_{0.8}Ge_{0.2}$ regions were hydrogenated by high-dose (5 x 10^{16} cm^{-2} @ 32 keV) implantation of H+ followed by a 10-minute forming-gas anneal at 275°C.
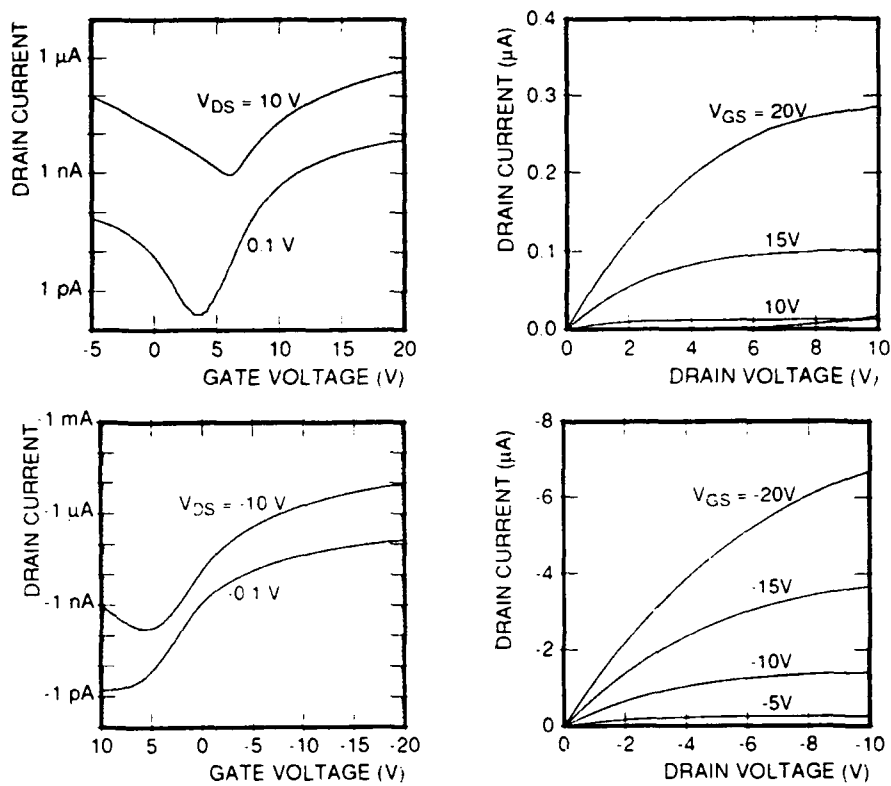
Figure 3   Measured TFT I-V characteristics (L = μm; W = 10μm).

Results and Discussion

Typical measured device characteristics are shown in Fig. 3, and a summary of TFT characteristics is listed in Table 1. Data for PMOS transistors fabricated using a 600°C process, from previous work [King2], has been included in Table 1, to show that poly-SiGe TFT performance is better when higher processing temperatures are used. The PMOS devices exhibit well-behaved characteristics, with fairly low (1 pA) leakage currents; however, the NMOS devices exhibit higher leakage currents, particularly under high drain bias. This is due to the fact that the undoped poly-$Si_{0.8}Ge_{0.2}$ channel material is n-type, so that the NMOS transistors are accumulation-mode devices.

	550°C process t_{oxide} = 50 nm		600°C process t_{oxide} = 108 nm
	NMOS V_{DS} = 0.1 V	PMOS V_{DS} = -0.1 V	PMOS V_{DS} = -1 V
Threshold voltage (V)	12	-8	-10
Peak channel mobility ($cm^2/V \cdot s$)	0.5	2	8
Sub-threshold slope (dec/V)	0.5	0.53	0.4
Leakage current (pA)	3	1	4
Maximum I_{ON}/I_{OFF}	6×10^3	2×10^5	5×10^5

Table 1   Summary of low-temperature-processed SiGe TFT characteristics (L=10μm; W=5μm). For comparison, data is shown for PMOS devices which were fabricated in Si0.75Ge0.25 films using a 600°C process [King2].

The TFT fabrication process parameters used in this work were not optimized for poly-$Si_{0.8}Ge_{0.2}$ devices. The results of our initial characterization work indicate that the temperature dependencies of film morphology, grain size, grain growth, and grain-boundary passivation (via hydrogenation) are very different for SiGe films compared to Si films. Presently, little is known about the trap states associated with dangling bonds and defects in poly-SiGe films. It is very likely that the energy distribution of these states is different from that in poly-Si films and changes with the Ge content in the film. Much more work needs to be done in order to understand the behavior of the grain boundaries and trap states as a function of deposition and subsequent process conditions. After a fundamental understanding of the traps in poly-SiGe films is attained, the passivation of these traps can be systematically studied. Since the device characteristics of polycrystalline TFTs are dominated by the electrical behavior of the traps located predominantly in grain boundaries of the thin-film channel region, effective grain-boundary passivation is essential for a poly-SiGe TFT technology to achieve high-performance devices.

It should be noted that it is difficult to achieve a fair comparison of the poly-SiGe TFTs in our initial investigation with current state-of-the-art poly-Si TFTs. While our present results do not show that the performance of poly-SiGe devices is superior to that of poly-Si devices, it can be expected that the overall performance of the poly-SiGe TFTs would improve greatly upon optimization of the crystallization and grain-boundary passivation processes, as is the case with poly-Si [Rodder]. In addition, a reduction in the channel poly-SiGe thickness should also improve device performance [Ohshima]. The results of Secondary Ion Mass Spectrometry (SIMS) analyses indicate that the poly-SiGe films have a relatively high oxygen content (mid 10^{19} cm^{-3}). Since the presence of oxygen in SiGe films has been shown to detrimentally affect the minority-carrier lifetime [Ghani], this may partially account for the low effective mobilities observed in our devices.

The most important result of this work is that NMOS and PMOS TFTs have been successfully fabricated in polycrystalline $Si_{0.8}Ge_{0.2}$ films without exceeding a process temperature of 550°C, using relatively short process times. This low-thermal-budget TFT technology could potentially allow integrated circuits to be

fabricated on large-area substrates like glass which cannot withstand higher-temperature and longer-time processing

References

[Ohwada]   J.-I. Ohwada et al., *IEEE Trans. Elec. Dev.,* **36**, 1923 (1989).

[King1]   T.-J. King et al., in *IEDM Tech. Dig.,* 253 (1990).

[King2]   T.-J. King et al., *IEEE Elect. Dev. Lett.,* **12**, 584 (1991).

[Pearsall]   T. P. Pearsall, *CRC Critical Reviews in Solid State and Materials Sciences,* **15**, 551 (1989).

[Rodder]   M. Rodder et al., *IEEE Elect. Dev. Lett.,* **EDL-6**, 570 (1985).

[Ohshima]   H. Ohshima et al., in *IEDM Tech. Dig.,* 157 (1989)

[Ghani]   T. Ghani et al., *Appl. Phys. Lett.,* **58**, 1317 (1991).

JSEP Supported Publications

1. T.-J. King, J. R. Pfiester and K. C. Saraswat, "A Variable-Worfunction Polycrystalline-$Si_{1-x}Ge_x$ Gate Material for Submicron CMOS Technologies," *IEEE Electron Device Letters,* **12**, 10, pp. 533-535, October 1991.

2. T.-J. King, K. C. Saraswat, and J. R. Pfiester, "PMOS Transistors in LPCVD Polycrystalline Silicon-Germanium Films," *IEEE Electron Device Letters,* **12**, 11, pp. 584-586, November 1991.

3. T.-J. King and K. C. Saraswat, "A Low-Temperature (<550°C) Silicon-Germanium MOS Thin-Film Transistor Technology for Large-Area Electronics," in *IEDM Technical Digest,* pp. 567-570, 1991.

4. T.-J. King and K. C. Saraswat, "Low-Temperature ($\leq$ 550°C) Fabrication of Poly-Si Thin-Film Transisotrs," submitted to *IEEE Electron Device Letters.*

5. M. Cao, T.-J. King and K. C. Saraswat, "Determination of the Density of Gap States in Undoped Polycrystalline Si and $Si_{0.8}Ge_{0.2}$ Films," submitted to *Applied Physics Letters.*

JSEP Supported Patents Filed

"A Low Temperature Germanium-Silicon on Insulator Thin-Film Transistor , and a Process for Forming the Same." (2 patents, with K. C. Saraswat.)

Unit: 8

TITLE: Signal Processing for Wideband Digital Protable Communications

PRINCIPAL INVESTIGATOR: T. H. Meng

GRADUATE STUDENT: C. Portmann

Scientific Objectives:

The objective of this research is to provide portable video on demand in a wireless environment. The main focus of the research is on the development of video compression algorithms with adaptive rate control and fault-tolerance capabilities and on the implementation of such algorithms with minimum power requirements.

Summary of Research:

A key component of this project is a low-power implementation strategy for the decoder design. We began with the design of a low-power Viterbi decoder, which has the capability of correcting corrupted data bits transmitted through a wireless link. This design also verifies the speed-power trade-off using voltage scaling. We also investigated error-resilient lossless encoding techniques to combat the bit errors that are not correctable by an error-correction code. The trade-offs among hardware (power) cost for error-correction, reduced compression ratios, and increased video quality as a result of added redundancy are quantified.

1. Low Power Design -A Viterbi-Decoder Chip

To demonstrate the power reduction that can be achieved through parallel processing, we designed a high-performance low-power Viterbi decoding chip for forward error-correction of transmitted data. To compare our design with some commercially available devices, we list two current products from Stanford Telecommunication Inc. [Stand, Bust]. It can be seen that our design outperforms these chips in both speed and power consumption.

Product	No. of States	Chip Area	Decoding Rate	Power Diss.
STI-2020	32,64	169	20Mbps	800mW
STI-2010	32	N/A	256Kbps	250mW
Ourchip @ 5V	32	62	140Mbps	1.8W
Ourchip @ 2.25V	32	62	56Mbps	110mW
Ourchip @ 2.25V	32	62	256Kbps	0.5mW

The error-correcting code specification of the decoder chip is 32-state, R=1/2, 8-level soft decision inputs, with generator polynomials G_1 = 111011 and G_2 = 110001. Coding gain as a function of probability of errors assuming additive white Gaussian noise is given as follows:

Error Prob	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Gain	3.6dB	4.3dB	4.8dB	S.OdB

At the decode rate of approximately 2Mbits per second necessary to support compressed video, the typical measured power dissipations for .the decoder as a function of supply voltage are listed below.

Voltage	5.0V	3.3V	2.25V
Power@ 2Mb/s	26mW	12mW	4mW

Although convolutional decoding using the Viterbi algorithm is one of the most complicated error-correction schemes, we have demonstrated that a high-speed, low-power implementation is feasible. The power consumed by the Viterbi decoder is minimal compared to what will be consumed by other functions performed at much higher rates (decompressed signals) in the portable terminal. Therefore, despite the algorithmic complexity of Viterbi decoding, we can use it in the portable terminal and benefit from its coding gain and error-correcting capabilities.

2. Error-Resilient Arithmetic Coding

The final stage of video coding is usually lossless compression, sometimes termed as "entropy coding". The application of lossless

compression is to eliminate as much redundancy as possible from the video data without adversely affecting any previous coding. We will consider two types of lossless compression techniques, the Huffman encoding and the arithmetic encoding, because of their high compression ratios and efficient hardware implementation. The advantages of a Huffman code are the relatively simple hardware complexity and better resilience against bit errors than an arithmetic code. The advantages of an arithmetic code are a higher compression ratio and the capability to adapt the codebook easily, but any error in the coded data results in corruption of all following data. This summary concentrates on the development of error-resilient arithmetic codes.

Two common methods are used to eliminate error propagation. One method uses resynchronization on the input symbols to the arithmetic coder. Every n input symbols, we place a unique start-of-interval marker and restart the encoding. We call this method input resynchronization. The other method is to use resynchronization on the output symbols of the arithmetic coder. Every n output symbols, we restart the encoding. We call this method output resynchronization. Input resynchronization can be applied to all encoding schemes with a constant overhead. We will discuss the use of output resynchronization for error-resilience.

Since variable numbers of input symbols can fit within n output symbols, output resynchronization may force encoding to be terminated a short while before the resynchronization interval is reached. Coding inefficiency at the interval boundaries is expected. This loss associated with output resynchronization can be quite substantial, as in the Fixari method described in [Teuh].
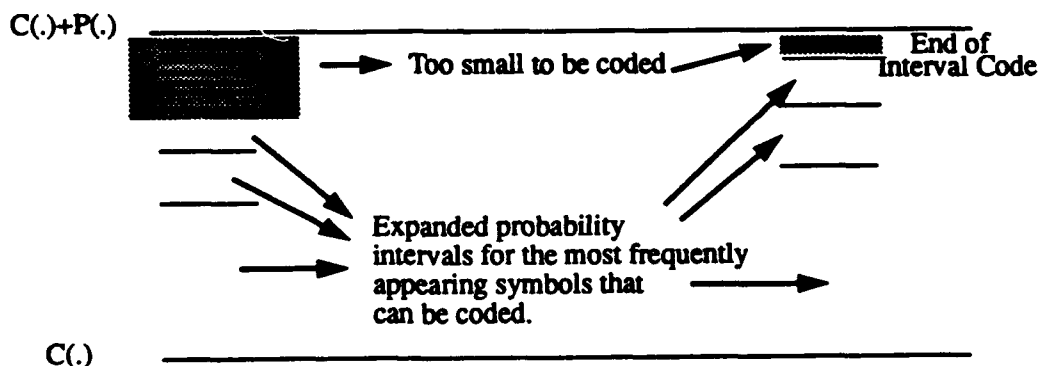

Figure 1   Reorder encoding prunes off the symbols that cannot be coded, thus increasing the interval sizes for those symbols that can be coded.

In order to recapture some of the lost efficiency, we develop a technique called "reorder" encoding. Symbol sequences are encoded such that the loss at the end of the output resynchronization interval is minimized. Using a permutation map, we can reorder the symbols such that the latest probability symbols occur first in the cumulative probability distribution. When we are close to the output resynchronization interval, we can discard all symbols from the symbol set that cannot be coded without exceeding the output resynchronization interval. With reordered cumulative probability, the most probable symbols will have a better chance of being encoded within the resynchronization interval, as shown in Fig. 1.

To demonstrate the trade-off between compression efficiency and error-resilience of each code, we first compare relative coding efficiency of the three arithmetic coding methods (the Fixari 32-bit resynchronization code, the reorder 32-bit resynchronization code, the arithmetic code without resynchronization) and the Huffman code. The simulation results show that the compression gain of using reorder resynchronization code over the Fixari code can be as large as 0.5 bit per symbol in compressed text files. The performance of the Huffman code is very close to the arithmetic (without output resynchronization) code in compression efficiency; both are about two-tenths of a bit more efficient than the reorder resynchronization code.

We then simulate the error resilience of the various coding techniques. A 10000-byte input file was encoded by the Fixari, reorder, and Huffman coders. A fixed codebook of symbol probabilities was generated and made available to the decoder, and a single error was placed in the encoded 1000 byte stream. The corrupted file was then decoded. The resulting numbers of correct symbols are shown in figure 2. Both the Fixari and the reorder coders perform better than the Huffman coder in the presence of one bit error. The reorder arithmetic coder is a reasonable compromise between compression efficiency and error-resilience. Furthermore, arithmetic coders can easily adapt their codebooks, a very desirable feature for video compression.
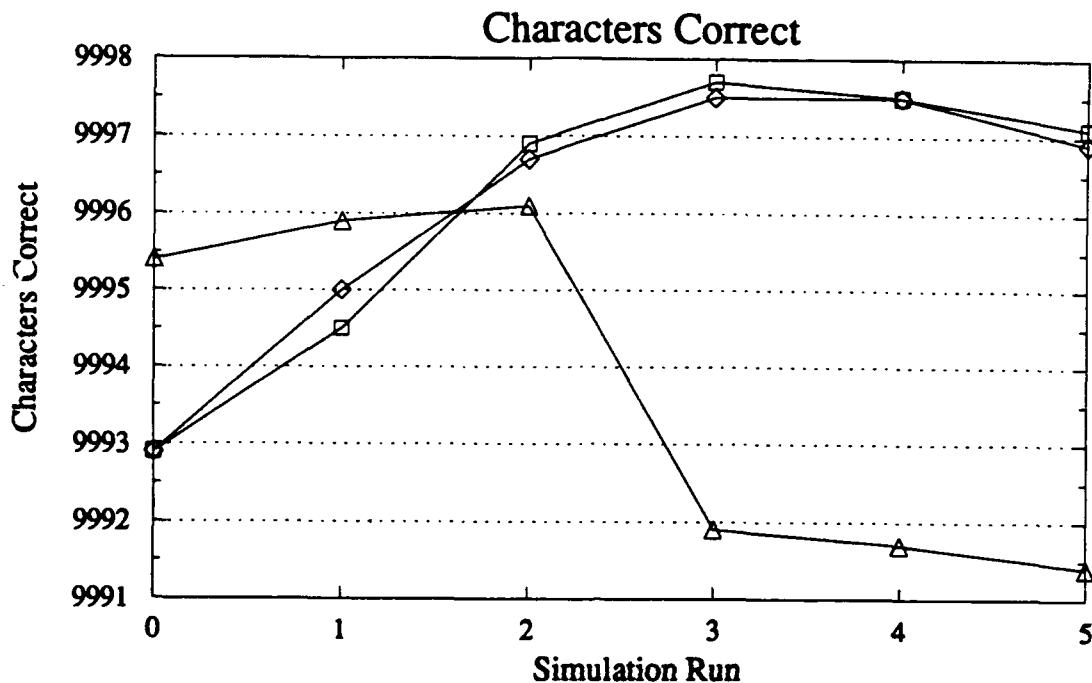
Characters Correct

Figure 2   The number of correct characters of the Fixari (square), the reorder (diamond), and the Huffman (triangle) coders as a result of one bit error.

3. Lossless Compression with Error-Correcting Codes

The goal of combining lossless compression and error-correcting codes is to minimize the transmitted data rate caused by the cascade of these two types of coding schemes, subject to the constraint of constant visual quality.

An error-correcting code effectively reduces the Eb/No ratio at which transmission at a given bit error rate (BER) can be achieved, thereby allowing reliable transmission in conditions under which successful transmission of uncoded data would be impossible. This reduction is referred to as "coding gain" at a given BER. The price of coding gain is an increase in data rates caused by the redundancy of the code. The error-correcting capabilities of a code are roughly proportional to the amount of redundancy that it adds to the data.

However, the proportionality constant varies greatly among different codes.

Because of the rate constraint, we consider only high-rate codes, specifically, the high-rate (R>3/4) convolutional codes. Convolutional codes have superb error-correcting capability but produce bursts of errors when the errors cannot be corrected. Thus interleaving of the input data to the convolution encoder is necessary to reduce burst errors in the decoded output.

A lossless compression scheme will be chosen based on two criteria: minimum compression ratio and maximum tolerance to error in the coded data. Four non-adaptive lossless compression schemes were compared by encoding fifty frames of a video sequence and simulating BERs of 10^{-4}, 10^{-5}, 10^{-6}, and 10^{-7} on each coded sequence. The sequences were then decoded and visually compared. An acceptable BER was defined as one in which errors in the decoded sequence were imperceptible to the viewer. The compression ratios, defined as the number of output bits divided by the number of input bits, were also noted. The results are listed below. Compression ratios for the corresponding adaptive schemes are also listed for comparison. It can be safely assumed that the adaptive schemes will have acceptable BERs equal to or worse than their non-adaptive counterparts, because any errors in the data will disturb the adaptation.

Compression Scheme	Compression Ratio	Acceptable BER
Fixed Fixari	0.73	10^{-6}
Fixed Reorder	0.71	10^{-6}
Fixed Arithmetic code	0.65	10^{-7}
Fixed Huffman	0.65	10^{-6}
Adaptive Fixari	0.72	$<10^{-6}$
Adaptive Reorder	0.69	$<10^{-6}$
Adaptive Arithmetic code	0.64	$<10^{-7}$

We briefly summarize the constraints in choosing a coding strategy. If the channel BER is around 10^{-4}, we require an error-correcting code that will reliably reduce the BER to a rate much lower than 10^{-6}, as the damage of one bit error in a data stream encoded by lossy source encoding will generally be more dramatic than the damage of one bit error in a data stream encoded losslessly. The overall tolerable BERs will depend upon both the lossy and the lossless encoder design. While lossless encoding using a Huffman code achieves

good compression ratios with tolerable error-resilience, the reorder arithmetic code may perform better with an intelligent strategy of adapting the codebook, which is currently under investigation.

Because the portable digital video system ultimately should operate with both a stationary and a mobile receiver, a coding scheme that is robust to the harsh mobile channel should be considered. Thus a convolutional code with highest constraint length possible will be employed. This will result in an overall reduction in data rate to 0.86 times the data rate out of a lossy encoder (0.65x1.33 = 0.86 using a R=3/4 convolutional code).

References

[Blac]   P. J. Black and T. H. Meng, "A 140Mbit/s, 32-State, Radix4 VIterbi Decoder", *1992 IEEE ISSCC Digest of Technical Papers*, February 1992.

[Bust]   H. A. Bustamante, I. Kang, C. Nguyen, and R.E. Peile, "Stanford Telecom VLSI Design of a Convolutiona! Decoder", *Proc. of MilCom 89*, Vol. 1, pp. 171-178, Oct 1989.

[Stan]   Stanford Telecommunications, "20Mbps Convolutional Encoder Vtterbi Decoder STEL-2020", Oct 1989.

[Teuh]   J. Teuhola and T. Raita, "Piecewise Arithmetic Coding," *Proceedings of the Data Compression Conference*, April 8-11, S n o w b i r d ,   U t a h ,   1 9 9 1 .

TITLE: Packet   Equalization

Principal   Investigator: J.   M.   Cioffi

**Graduate   Students: N.   Al-Dhamir,   H.   Bims
and   S.   K.   Wilson**

Scientific   Objective
The purpose of our study has been identification of reliable and efficient methods for data transmission and detection in a mobile environment.   Particular emphasis has been given to situations where data rates are high, the number of users is large, and multipath fading can dominate performance.

Summary   of   Research
The focus of our study has been in two main areas: reduced-complexity channel-based equalization methods and interference rejection using code division multiple access techniques.   In the first area, we were able to derive new algorithms for the computation of the decision feedback equalizer settings from channel estimates.   These algorithms have been shown to be much more efficient than currently used adaptive techniques such as recursive least squares (RLS).   In the second area, we have examined equalization techniques for Code Division Multiple Access (CDMA) Systems.   Spread-Spectrum CDMA systems are competing with Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) systems for use in commercial mobile communications. Methods to increase the SNR in CDMA systems have so far focused on removing noise due to other users in a non-dispersive channel.   However, in a wireless environment the channel is dispersive due to multipath distortion between the transmitter and receiver.   We have investigated methods to reduce ISI and interuser interference IUI in a dispersive multipath channel.

1.   Efficient   channel-based   equalization
The minimum mean-square-error decision feedback equalizer (MMSE-DFE) is a long-established ISI-mitigating technique for data transmission on multipath dispersive mobile channels.   The optimum MMSE-DFE filter settings can be calculated recursively using the

least mean square (LMS) or recursive least squares (RLS) algorithms, which we call "adaptive equalization". We can also indirectly compute the MMSE-DFE settings from explicit estimates of the channel pulse response, which we call "channel-based equalization". Over the last two decades, most research effort was directed towards adaptive equalization. It has not been until recently that more attention has been given to channel-based equalization, because of the performance improvements they achieve [Ziegler91]. On the other hand, the main drawback of channel-based equalization methods for DFE adaptation has been their high computation requirements.

1.1 Model and Assumptions

Our interest is in block-oriented TDAM systems (which have already been chosen for the new mobile-radio standards in the U.S. and Europe [EIA/TIA89], where data is transmitted in fixed-length packets. Within each packet, a known "training" sequence is embedded to aid in estimating the channel pulse response. This estimated response is then used to detect the unknown information symbols that comprise the rest of the block using a minimum mean-square-error decision feedback equalizer (MMSE-DFE).
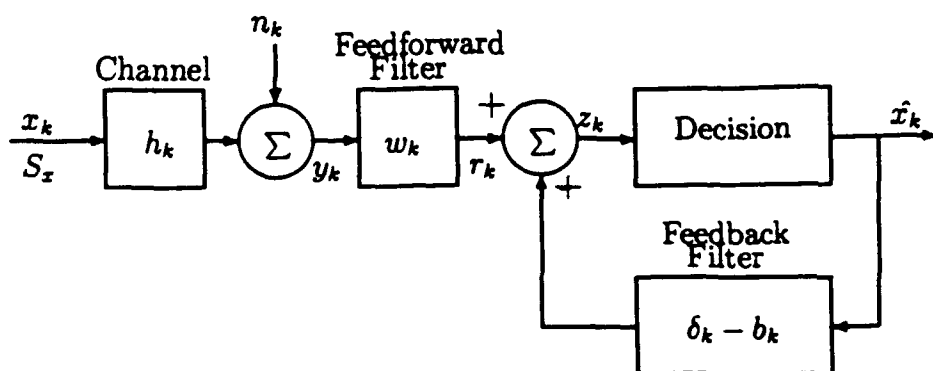

Figure 1 Block diagram of the MMSE-DFE

The basic structure of the discrete-time MMSE-DFE is shown in Fig. 1. Assuming that the channel is time-invariant over a block of (N) output symbols, then the input-output relation can be cast in matrix form as follows:

$$
\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-N+1} \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & \cdots & h_\nu & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_\nu & 0 & \cdots \\ \vdots & & & & & & \vdots \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_\nu \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-N-\nu+1} \end{bmatrix} + \begin{bmatrix} n_k \\ n_{k-1} \\ \vdots \\ n_{k-N+1} \end{bmatrix} \qquad (1)
$$

Within each block, we make the following assumptions

- X_k is a complex, independent and identically-distributed data sequence with energy S_x per complex dimension. The noise is white Gaussian with a power spectral density of (N_o) per complex dimension.

- Previous decisions are assumed to be correct.

- The filter $b(D) = 1 + b_1 D + b_2 D^2 + \ldots + b_\nu D^\nu$ is monic and has $\nu + 1$ taps (same number of taps as the channel pulse response.

1.2 Finding the Optimum MMSE-DFE Settings
By considering the following "Cholesky" (Lower-Diagonal-Upper, * denotes the conjugate transpose) factorization:

$$
R = \frac{1}{SNR} I_{N+\nu} + H^*H = LDL^* \qquad (2)
$$

where, **H** is the Toeplitz channel matrix given in (1) and SNR = S_x/n_o, we derived in the following expressions for the optimum DFE tap settings:

$$
\begin{bmatrix} 0_{(N-1)\times 1} \\ b \end{bmatrix} = Le_N \qquad (3)
$$

$$
w^* = d_{N-1}^{-1} e_N^* L^{-1} H^* \qquad (4)
$$

where, d_{N-1} is the Nth diagonal element of **D** and e_N is the Nth unit vector (has a one in the Nth position and zeros everywhere else.

1.3 Fast Algorithm for Computation of MMSE-DFE Settings

Using the generating function algorithms for fast Cholesky factorization of structured matrices [Lev-Ari86], we derived in [Al-Dhahir91] the following fast algorithms for the computation of the filter b(D).

Algorithm (Symbol-spaced case):
Initial Condition:

$$G_0(D) = G(D) = \left[\frac{1}{\sqrt{SNR}} \quad h^*(D^*) \right]$$

Recursions:
For $i = 0, 1, \ldots, (N-1)$:

$$d_i = |G_i(0)|^2 \tag{5}$$

$$G_i(0) = d_i^{\frac{1}{2}} \left[\alpha_i \quad \beta_i \right] \tag{6}$$

$$l_i(D) = D^i G_i(D) G_i^*(0) d_i^{-1} \tag{7}$$

$$DG_{i+1}(D) = G_i(D) \begin{bmatrix} \alpha_i D & -\beta_i \\ \beta_i^* D & \alpha_i \end{bmatrix} \tag{8}$$

Where $h^*(D^*) = \sum_{i=0}^{\nu} h_i^* D^i$. It follows that b(D) = l$_{N-1}$(D). We generalized these algorithms in [Al-Dhahir91] to the case of a fractionally-spaced feedforward filter. Once the lower triangular Cholesky factor **L** has been computed, the feedforward filter **w*** can be efficiently computed by **Back Substitution** as described in [Al-Dhahir91].

1.4 Computational Complexity Evaluation

A detailed computational complexity evaluation of our algorithms appears in [Al-Dhahir91]. Because of space limitations, we demonstrate the high efficiency of our method by an example. Consider packet data transmission on the proposed U.S. digital cellular radio standard IS-54 [EIA/TIA89], which sets the terminal data rate at 24 kbaud/sec. Figure 2 plots the total MIPS estimate vs. different choices for the number of feedforward (N) and feedback ℓ, taps, for a DFE coefficient update rate K = 2. This estimate is considerably less than current processing rates attained with adaptive RLS, and can be further reduced by increasing K to 3 or even 5 in some situations without significant performance degradation [Ziegler91]. Similar results for the fractionally-spaced case are given in [Al-Dhahir91].
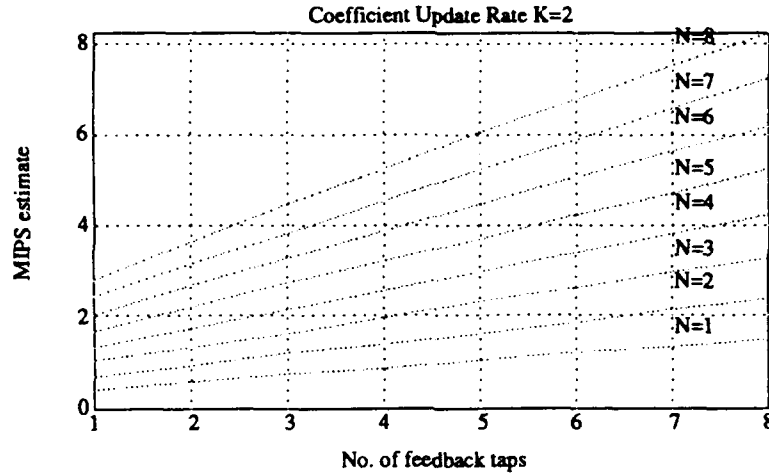
Figure 2 MIPS estimate for the tracking of a T-spaced MMSE-DFE

2. Equalization techniques for high data rate CDMA

We investigated equalization techniques for reducing ISI and IUI and CDMA systems. We found that at high data rates, the ISI is about 10% of the signal and the major source of noise was IUI. Verdu [Verdu86] studied the optimum maximum-likelihood detector for CDMA systems with a non-dispersive channel. This method, though optimal for CDMA systems with both ISI and IUI, has high complexity. We examined suboptimal techniques that reduce complexity while sacrificing little SNR. We can treat a CDMA system as a multi-dimensional system with both ISI and IUI. The received signal can be modeled as:

$$ y = \left(\begin{array}{cccc} H_1 & H_2 & \ldots & H_K \end{array} \right) \begin{pmatrix} C_1 & 0 & 0 & \ldots & 0 \\ 0 & C_2 & 0 & \ldots & 0 \\ & & \vdots & & \\ 0 & 0 & \ldots & 0 & C_K \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix} + n $$

where x_k is the vector of data, C_k is the matrix of spreading sequences and H_k is the channel matrix for the kth user and n is additive white noise. We assume that the spreading sequences are *N* chips long. This is illustrated in Fig. 3.
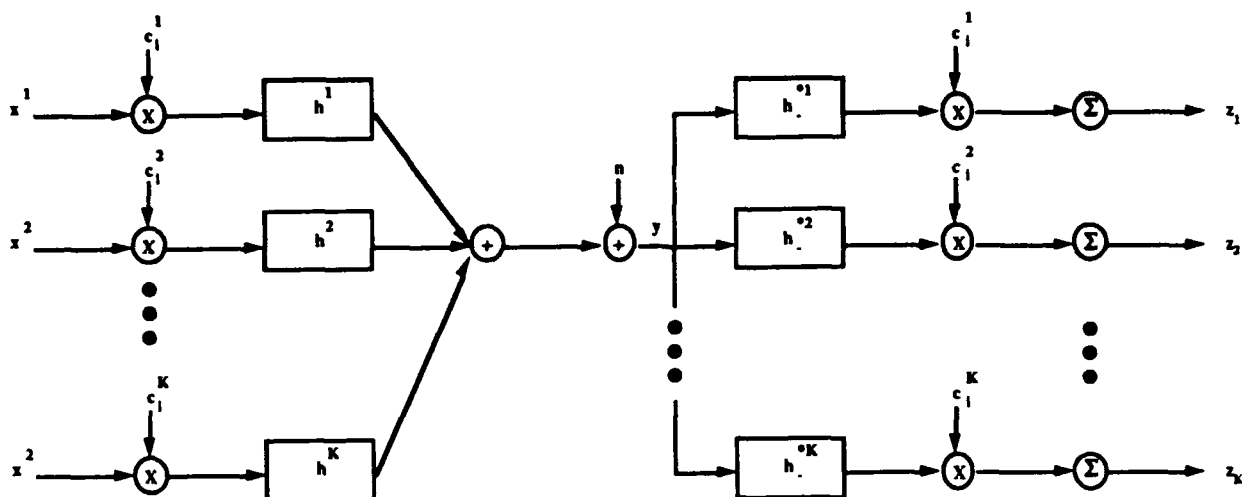
71

Figure 3 Conventional Single-user spread-spectrum detector.

2.1 Multi-Dimensional Equalizers for CDMA systems in Multi-path Environments

In our research, we evaluated the performance of Multi-dimensional Equalizers vs. a RAKE receiver. We evaluated multi-dimensional Zero-Forcing Equalizers (ZFE), Minimum Mean Square Error Linear Equalizers (MMSE-LE) and Decision Feedback Equalizers. The multi-dimensional ZFE equalizer is a generalization of the Linear Multiuser Detector developed by Lupas and Verdu [Lupas90]. We found that we can equalize at the bit rate rather than the chip rate. This is because the despread data at the bit rate for each user is a sufficient statistic for the data at the chip rate. This allows us to equalize at a much lower sufficient statistic for the data at the chip rate. This allows us to equalize at a much lower bandwidth, ensuring a faster, less complex implementation. We equalized the data using 3 users, and treating the other users in the CDMA system as additive white Gaussian noise. We found that 3 dimensional equalization in general provided about 7 dB improvement over a conventional single user RAKE receiver. We tested 3 types of equalizers at different data rates. In Fig. 4, we see that for a chip rate of 40 chips/bit, and a data rate of approximately 50 Mb/s, there is about a 2dB improvement of a 3D DFE over a 3D MMSE-LE and 3D ZFE. The 3D MMSE-LE and 3D-ZFE in turn have about a 7 dB improvement over the conventional RAKE receivers. At a slightly lower data rate of approximately 10 Mb/s, there is no large difference
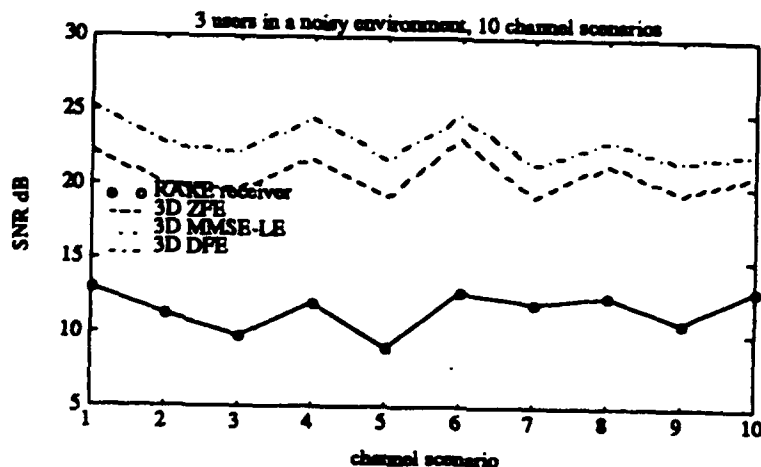
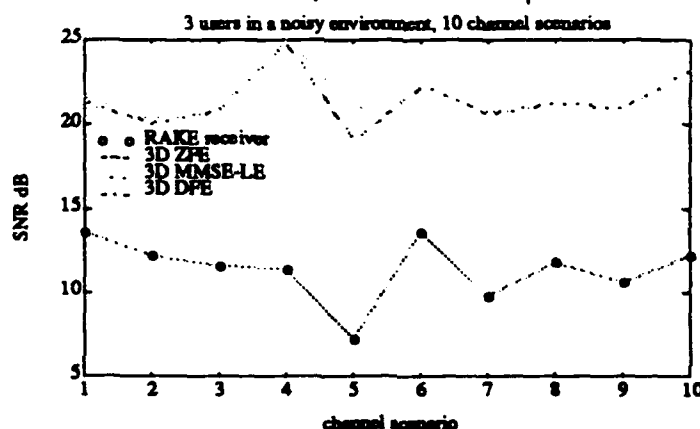Figure 4 High data rate comparison of Equalizers

Figure 5 Low data rate comparison of Equalizers

between the 3D DFE, 3D MMSE-LE and 3D ZFE; but there is again a 7dB improvement over the conventional RAKE receiver. This is illustrated in Fig. 5.

This implies that we need an excessively large data rate to warrant the additional complexity of a multi-dimensional DFE over a ZFE equalizer. This means that for relatively low complexity, we can invert the effect of the cross-channels and the additional ISI. The reason it takes very high data rates for a DFE to show significant improvement over the 3D ZFE, is that the ISI must be about 50% of the main sampling period to have significant noise enhancement from a ZFE. The effective channel for a CDMA system after spreading, matching and despreading in a multipath environment, is relatively non-dispersive unless we have extremely high data rates and bandwidth. As the chip sequences vary for each bit, the 3D ZFE will need to be updated for each bit. Fast methods for this need to be investigated.

73

References

[Al-Dhahir91]   N. Al-Dhahir and J. Cioffi, "Fast computation of channel-estimage based equalizers in packet data transmission," submitted to *IEEE Trans. on Signal Processing*, November 1991.

[EIA/TIA89]   EIA/TIA project number 221.5 dual-mode mobile station-base stations compatibility standard (IS-54), December 1989.

[H. Lev-Ari86]   H. Lev-Ari and T. Kailath, "Triangular factorization of structures hermitian matrices," *I. Schur Methods in Operator Theory and Signal Processing*, **18**, 301-324, 1986.

[Lupas90]   R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. on Information Theory*, **35**,(1), 123-126, January 1989.

[Ziegler91]   R. Ziegler, N. Al-Dhahir and J. Cioffi, "Non-recursive adaptive decision-feedback equalization from channel estimates, submitted to *IEEE Trans. on Communications*, July 1991.

JSEP Supported References

[Al-Dhahir91]   N. Al-Dhahir and J. Cioffi, "Fast computation of channel-estimage based equalizers in packet data transmission," submitted to *IEEE Trans. on Signal Processing*, November 1991.

[Bims]   H. Bims and J. Cioffi, "Trellis Coding for Partial-Response CPM Using an MSK-type Modulation," in *Proceedings IEEE International Symposium on Information Theory*, July 1991.

[Ziegler]   R. Ziegler, N. Al-Dhahir and J. Cioffi, "Non-recursive adaptive decision-feedback equalization from channel estimates, submitted to *IEEE Trans. on Communications*, July 1991.

TITLE: Fast Computing with Neural Networks

Principal Investigator: T. Kailath

Graduate Students: K.-Y. Siu and G. Xu

Scientific Objectives:

The aim of this project has been to examine new methods of computing using artificial neural networks. During the past year we refined our earlier results on a new class of fast arithmetic circuits based on the neural network model, and began some new work on the complexity analysis of neural networks with continuous output units.

Summary of Research:

A. A New Class of Fast Arithmetic Circuits based on the Neural Network model

The basic processing element of our neural networks is a *linear threshold element* (LTE). It is well known that LTEs can simulate digital logic gates such as AND, OR, NOT. Therefore, we can perform digital arithmetic such as additions and multiplications using neural networks. The basic problem is that a multiplication cannot be done in a digital circuit with a <u>fixed</u> delay. One of the basic assumptions in the study of neural networks is that we allow the LTE to take on an unbounded number of inputs, i.e., we allow *unbounded fan-in*. Perhaps this feature of LTE can give rise to a significant increase in computing power? In fact, our earlier results showed that for example, the product of two n-bit numbers could be computed with <u>4 unit delays</u> with neural networks. Moreover, the weights of our networks only require $O(\log n)$-bit accuracy. Recently, we have improved and extended our results to cover more complicated functions and have shown that the division of two integers and the computation of the power of an integer can also be done with <u>4 unit delays</u>, and that multiple products can be computed with <u>5 unit delays</u>.

We also examined the tradeoffs between the speed and the amount of hardware when the fan-in of the gates is *large but restricted*. It is known that the potential speed-up in performing multiplications by increasing the fan-in of the AND-OR circuit is minimal. In fact, even if the fan-in is allowed to be arbitrarily large, the best possible propagation delay in an AND-OR circuit for multiplication of 2 n-bit integers has been shown to be $\Omega(\log$

$n/\log\log n$). However, with neural networks, we can design a multiplier that speeds up asymptotically with respect to the maximum allowable fan-in. In particular, we have shown that a neural network can be constructed for every $d > 1$ with $O(2^{-d/2}n^2m^{1/(2d-1)}\sqrt{\log m})$ connections and $O(d\log n/\log m)$ number of layers, at most m fan-in. For example, when $n = 64$ and fan-in $m = 16$, our design will yield a threshold circuit of about 4,000 connections and 8 propagation delays, whereas if we use AND/OR logic gates, then the best known circuit with approximately the same number of connections requires 30 delays.

B. Complexity Analysis of Neural Networks with Continuous Output Units

It is commonly agreed that the computational power of neural networks comes from the fact that each basic processing element computes a nonlinear function of its inputs. The interconnection of these nonlinear elements can yield solutions to highly complex and nonlinear problems. On the other hand, because of the nonlinear features, it is very difficult to study the fundamental limitations and capabilities of neural networks.

Most previous work on analyzing the complexity of neural networks has been focused on the threshold circuit model, in which each basic processing unit is a threshold element. Recently, it was shown that in some cases, networks of continuous-output units, such as the sigmoidal elements, are far more powerful than networks of linear threshold elements with binary outputs. We have extended our earlier work on threshold circuits to a more general class of neural networks. We employ harmonic analysis and rational approximation theory to characterize the class of Boolean functions whose complexity is almost the same among various models of neural network]s with feedforward structures. As a consequence of this characterization, for example, we prove that any depth-$(d + 1)$ network of sigmoidal units computing the parity function of 71 variables must have $\Omega(dn^{1/d-e})$ units, for any fixed $\varepsilon > 0$. This lower bound is almost tight since we can compute the parity function with $O(dn^{1/d})$ sigmoidal units in a depth-$(d + 1)$ network. Our techniques also generalize to networks whose elements can be approximated by piecewise low degree rational functions. These almost tight bounds are the first known complexity results on the size of neural networks computing Boolean functions with continuous-output elements and with depth more than two. Since essentially the same bounds can be achieved by linear threshold elements, these results

imply that when computing highly oscillating discrete functions such as the parity, using continuous-output elements such as the sigmoidal units would not give us advantages over binary-output linear threshold elements.

C.   We continued with the publications process on several of our earlier papers, especially those on sensor array processing.

JSEP SUPPORTED PUBLICATIONS
Journal Papers

I.   D. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least-squares adaptive filtering," *IEEE Trans. Acoust. Speech Signal Process.*, **39**(1):92-114, Jan 1991.

2.   A. Dembo, 0. Farotimi and T. Kailath, "High Order Absolutely Stable Neural Networks", *IEEE Trans. Circuits and Systems*, **38**(1):57-65, Jan. 1991.

3.   0. Farotimi, A. Dembo and T. Kailath, "A General Weight Matrix Formulation Using Optimal Control," *IEEE Trans. Neural Networks*, 2(3):378-394, May 1991.

4.   B. Ottersten, M. Viberg and T. Kailath, "Performance Analysis of the Total Least Squares ESPRIT Algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, **39**(5):1122-1135, May 1991.

5.   M. Viberg and B. Ottersten, "Sensor Array Processing Based on Subspace Fitting," *IEEE Trans. Acoust. Speech Signal Process.*, **39**(5):1110-1121, May 1991.

6.   K-Y. Siu and J. Bruck, "On The Power of Threshold Circuits with Small Weights," *SIAM J. Discrete Math.*, **4**(3):423-435, Aug. 1991.

7.   A. Dembo, T. M. Cover and J. Thomas, "Information Theoretic Inequalities," *IEEE Trans. Inform. Theory,* **37**(6):1501-1518, November 1991.

8.   M. Viberg, B. Ottersten, and T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting," *IEEE Trans. Acoust. Speech Signal Process.*, **39**(5):1110-1121, May 1991.

9.   A. Dembo, T. M. Cover and J. Thomas, "Information Theoretic Inequalities," *IEEE Trans. Inform. Theory,* **37**(6):1501-1518, November 1991.

10.   M. Viberg, B. Ottersten and T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting," *IEEE Trans. Acoust. Speech Signal Process.*, **39**(11):2436-2449. Nov. 1991.

11. K-Y. Siu, V. Roychowdhury and T. Kailath, "Depth-Size Tradeoffs for Neural Computation," *IEEE Trans. Comput.*, **40**(12):1402-1412, Dec. 1991.
12. L. Chisci, H. Lev-Ari, D. T. M. Slock and T. Kailath, "Fast parallel self-tuning controllers," *Internat. J. Control*, **54**(6):1353-1384, Dec. 1991.
13. M. Viberg, B. Ottersten and T. Kailath, "Subspace Based Detection for Linear Structural Relations," *J. Combin. Inform. System Sci.*, 16(2&3):170-189,1991.

Published Conference Papers

14. G. Xu and T. Kailath, "A Fast Algorithm for Signal Subspace Decomposition and Its Performance Analysis," pages 3069-3072, Toronto, Canada, May 1991, 1991 Intern'l. Conf. Acoustics, Speech and Signal Processing.
15. K-Y. Siu, V. Roychowdhury and T. Kailath, "Computing With Almost Optimal Size Threshold Circuits," Budapest, Hungary, June 1991. Intern'l. Symp. on Inform. Thy.
16. G. Xu, R. H. Roy, and T. Kailath, "Strong consistency of FSD detection schemes," Pacific Grove, CA, Nov 4-6 1991, 25th Annual Asilomar Conf. on Signals, Systems and Computers.
17. G. Xu and T. Kailath, "Fast signal subspace decomposition based on data matrices," Pacific Grove, CA, Nov 4-6 1991, 25th Annual Asilomar Conf. on Signals, Systems and Computers.
18. K-Y. Siu, V. Roychowdhury and T. Kailath, "Circuit complexity for neural computation," Pacific Grove, CA, Nov 4-6 1991, 25th Annual Asilomar Conf. on Signals, Systems and Computers.

Unit: 11

TITLE: Research in Image Compression, Data Compression and Network Information Flow

Principal Investigator: T. M. Cover

Graduate Students: P. Algoet, J. Roche, A. Nobel and E. Erkip

1 Scientific Objectives

The objective of our work is to apply the techniques of information theory to the problems of image compression and image transmission, distributed data compression and storage, and network information flow. This work has resulted in six supported papers, three conference presentations, two Ph.D. theses in the past year.

2 Summary of Research

2.1 Introduction

We report substantial results in several areas. In [Barron and Cover] we proposed a new technique for density estimation which could be used for adaptive data compression. In [Equitz] we derived the conditions which allow successive refinement to be rate-distortion optimal at each stage. This work has large implications for the progressive coding of images and other data. In [Nobel] we extended the classical law of large numbers from a single function to an entire class of functions. This powerful theoretical result will allow a more rigorous understanding of why many inductive procedures such as pattern recognition, vector quantization, and neural networks behave as they do.

Furthermore, in [Ekroot and Cover a] and [Ekroot and Cover b], we derived the entropy both for Markov trajectories and for randomly stopped sequences. This work provides a clear understanding of the descriptive complexity of the many real world processes which can be modeled as Markov chains or as randomly stopped sequences. We also showed in [Cover] that Huffman codes are optimal in the short term as well as in the long term. This result shows that there is no need to search for a better method for source coding even if the message being compressed is short.

In [Roche] we report strong preliminary results for fault tolerant distribution of information among several sites which are failure prone. Our work shows that many sites can be given only partial information with no site being more critical to the recovery of the information than any other site. In [Pombra] and [Pombra and Cover], we report new results bounding the capacity increase possible by using feedback on single user and multiple access channels with non-white additive Gaussian noise. Finally, in the invited paper [Dembo et al], we reviewed many of the most important inequalities in information theory and examined their relationships to classical inequalities in mathematics.

2.2 Minimum Complexity Density Estimation

The minimum complexity, or minimum description-length, criterion developed by Kolmogorov, Rissanen, Wallace, Sorkin, and others is shown in [Barron and Cover] to lead to consistent probability density estimators. These density estimators are defined to achieve the best compromise between likelihood and simplicity. This serves as an extension of the MDL method from parametric density estimation to the nonparametric case. A related issue explored in this article is the compromise between the accuracy of approximations and complexity relative to the sample size. Barron and Cover propose an index of resolvability which bounds the statistical accuracy of the density estimators as well as the information theoretic redundancy.

2.3 Sequential Refinement

The successive refinement of information consists of first approximating data using a few bits of information, then iteratively improving the approximation as more and more information is supplied. An ideal system would provide a description which is rate-distortion optimal at each stage. In [Equitz] it is shown that such an ideal solution is possible if and only if the individual solutions of the rate-distortion problems at each stage can be written as a Markov chain. This implies in particular that tree structured descriptions are optimal if and only if the rate-distortion problem is successively refinable.

Equitz further shows that ideal successive refinement is possible for all finite alphabet signals with Hamming distortion, for Gaussian signals with squared error distortion, and for Laplacian signals with absolute-error distortion. However, Equitz uses a simple counterexample with absolute error distortion and a

symmetric source distribution to show that successive refinement is not always achievable.

2.4 Uniform Laws of Averages

In [Nobel], the classical law of large numbers is extended from a single real valued function to a collection of such functions. This extension provides a methodology with which the consistency of inductive procedures for many statistical estimation problems including pattern recognition, vector quantization, and the training of neural networks. Nobel uses closure properties that hold when the underlying observation are independent and identically distributed to extend uniform laws of averages to empirical quantities such as maxima or minima that cannot be evaluated in terms of averages. Nobel makes a further extension to an analogous law for weakly dependent (absolutely regular) observations.

2.5 Entropy Derivations

Entropy derivations for Markov trajectories [Ekroot and Cover a] and for randomly stopped sequences [Ekroot and Cover b] are derived. The entropy of a Markov trajectory is essentially the descriptive complexity of the path through which a system evolved from one state into another state.

A trajectory T_{ij} of the Markov chain is a path with initial state i, final state j, and no intervening states equal to j. Ekroot and Cover find that the entropy $H(T_{ii})$ of the random trajectory originating and terminating in state i is given by $H(\mathcal{X})/\mu_i$, where μ_i is the stationary probability of state i and $H(\mathcal{X})$ is the entropy rate of the stationary Markov process. Thus the entropy of the random trajectory T_{ii} is the product of the expected number of steps $1/\mu_i$ to return to state i and the entropy rate $H(\mathcal{X})$ per step for the stationary Markov chain.

Ekroot and Cover find a general closed form solution for the trajectory entropies $H(T_{ii})$ as

$$H = K_1 - K_2 + H_\Delta$$

where H is the matrix of trajectory entropies with $H_{ij}=H(T_{ij})$ and the matrices K_1, K_2 and H_Δ are matrices which depend on the stationary probabilities, the transition probabilities, and the entropy rate of the Markov process.

In closely related work on randomly stopped sequences, Ekroot and Cover prove that a Wald-like equation holds when the sequence consists of independent identically distributed discrete random variables X_1, X_2 ... with a nonanticipating stopping time N. Specifically, they showed that

$$H(X^N)=(EN)H(X_1) + H(N|X^\infty)$$

where X^N denotes the randomly stopped sequence. Thus the randomness in the stopped sequence is the expected number of "calls" for X times the entropy per call plus the residual randomness in the stopping time conditioned on the unstopped sequence X^∞.

2.6 Short Term Optimality of Huffman Codes

It is well known that the Huffman codes are optimal in the long term. That is, they minimize the expected description length. In [Cover], the performance of Huffman and Shannon codes in the short term is examined. Let X be a discrete random variable drawn according to a probability mass function $p(x)$. If $p(x)$ is dyadic, i.e., $\log(1/p(x))$ is an integer for each x, then the Shannon code length assignment $l(x) = \log(1/p(x))$ is also the Huffman code length assignment. In this case it is shown that

$$Pr\{ l(X)< l'(X)\} > Pr\{ l(X) > l'(X)\}$$

where $l'(x)$ is any other uniquely decodable assignment. This indicates that l is optimal in the short run since it is more likely to produce the smaller codeword. In general, if $p(x)$ is not dyadic, we have the following weaker result:

$$Pr\{ l(X)< l'(X)+1\} > Pr\{ l(X) > l'(X)+1\}.$$

2.7 Distributed Information Storage

Using techniques from Galois field theory and network flow theory, in [Roche], a Ph.D. thesis currently in preparation, we show how to distribute information among several different sites with a minimum of redundancy so that no data is lost even if several sites become inaccessible.

A simple example of the powerful results of this work involves a sequence of t bits of which t bits are distributed to each of K sites in suc' a way that if all but two of the sites fail, the original sequence can be recovered. How many different sites

can be used? Clearly there are at least three since we could store the first half of the sequence at the first site, the second half at the second site, and the exclusive or of the first half and the second half at the third site. Using combinatorics, Roche shows that $K \leq 2^t + 1$. Using Galois field theory, Roche surprisingly shows this bound to be *achievable*.

Having found theoretical limits on the amount of information that can be reliably stored in various networks, attention is focused on the configuration of disks that is widely used in practice. Roche demonstrates how to reduce the time spent updating parity-check disks when data disks are continually being modified.

2.8 Multiple Access Channels with Gaussian Noise

Gaussian channels model many practical channels including radio and satellite links. It is well known that feedback does not increase the capacity of a single user additive white Gaussian noise channel. In [Pombra] and [Pombra and Cover] we examine the degree to which feedback can be used to increase the capacity for the single user channel with non-white Gaussian noise and the multi-user or multiple access channel with non-white Gaussian noise.

For the single user channel with non-white Gaussian additive noise, we prove the following two inequalities which relate the nonfeedback capacity C and the feedback capacity C_{FB}: $C \leq C_{FB} \leq 2C$ and $C \leq C_{FB} \leq C + 1/2$. Thus feedback will at most double the capacity and feedback will at most increase capacity by half a bit.

For the multiple access channel with non-white Gaussian additive noise, we show that feedback increases capacity by at most *1/2 log (μ+1)* bits where μ is the number of users. We further show that the factor of two bound from the single user case still holds even for the multiple access channel. This paper [Pombra and Cover] is currently under review by the IEEE Transactions on Information Theory.

2.9 Information Theoretic Inequalities

As demonstrated in the previous section, major results in information theory of practical significance to communications often take the form of inequalities. In an invited paper [Dembo et al], we review many of the fundamental inequalities in information theory and explore their rela*   * *   *   to inequalities in other branches of mathematics. Th*   *avor improves and expands understanding of the fo** * *   *** of information theory.

2.10 Conclusion

We report substantial results in several areas including density estimation, sequential refinement, uniform laws of averages, the entropy of Markov trajectories and randomly stopped sequences, the short term optimality of Huffman codes, distributed information storage, feedback for Gaussian channels, and information theoretic inequalities. These results have practical implications for adaptive data compression, progressive coding, vector quantization, neural networks, and the fault tolerant distribution of information among various sites.

JSEP Supported Publications

[Cover] T. Cover, "On the Competitive Optimality of Huffman Codes," *IEEE Trans. on Info. Theory*, 37(1): 172-174, January, 1991.

[Dembo et al] A. Dembo, T. Cover, and J. Thomas, "Information Theoretic Inequalities," *IEEE Trans. on Info. Theory*, 37(6): 1501-1518, November 1991.

[Ekroot and Cover a] L. Ekroot and T. Cover, "The Entropy of Markov Trajectories," *Stanford Statistics Department Technical Report*, to appear in 1992, under review by *IEEE Trans. on Info. Theory*.

[Ekroot and Cover b] L. Ekroot and T. Cover, "The Entropy of a Randomly Stopped Sequence," *IEEE Trans. on Info. Theory*, 37(6): 1641-1644, November 1991.

[Equitz and Cover] W. Equitz and T. Cover, "Successive Refinement of Information," *IEEE Trans. on Info. Theory*, 37(2): 269-275, March 1991.

[Pombra and Cover] "Non-White Gaussian Multiple Access Channels with Feedback," submitted in 1991 to *IEEE Trans. on Info. Theory*.

JSEP Supported Ph.D. Theses

[Ekroot] L. Ekroot, "Process Entropy," September, 1991

[Pombra] S. Pombra, "Non-white Gaussian Channels with Feedback," September, 1991.